Dynamical Determinants and their Applications

by

Philip Felton

Thesis

Submitted to The University of Warwick

for the degree of

Doctor of Philosophy

Mathematics Institute

December, 2010

Table of Contents

Acknowledgments

I am indebted to my supervisor Professor Mark Pollicott, who has offered an abundance of invaluable time and support during my time at the University of Warwick. I am privileged to be his student.

I would like to thank my friends and family for their support and encouragement, and my employer Mercedes Benz Grand Prix Ltd for their flexibility while I have been writing this thesis.

I would particularly like to thank my daughter Willow for her patience and understanding of my time pressures.

Declaration

I declare that the work in this thesis is, to the best of my knowledge, original and my own work, except where otherwise indicated, cited, or commonly known. This work has not been submitted for any other degree.

The chapter on entropy rates of hidden Markov processes is based on earlier ideas of Pollicott (44), published as a survey in conference proceedings. However, the contribution of this chapter is to give a rigorous investigation of the details.

Abstract

This thesis is concerned with situations where we can define trace-class transfer operators, and extract useful information from their determinants.

The first topic is on Lyapunov exponents of random products of matrices. We obtain a new expression for the Lyapunov exponent of a continuous family of matrices, and a slightly different version of existing work for the discrete case.

The second topic explores possibilities of using similar theory to approximate eigenfunctions of the Laplacian for surfaces of constant negative curvature.

The third topic gives a variety of approximations of Mahler measures, which occur in many different areas of mathematics, by manipulating the integrals into a form that can be numerically integrated using work of Pollicott and Jenkinson.

The final topic of the thesis works out the details of earlier ideas of Pollicott, to give a method for the numerical approximation of entropy rates of hidden Markov processes.

Chapter 1

Introduction

The general theme of this work is to examine a variety of settings in which transfer operators can be defined and be shown to be trace class, so that we may then define the associated dynamical zeta function, enabling us to finally say something useful about appropriate characteristics in that setting. Previous examples include the following:

- Jenkinson and Pollicott (22; 23) approximate the Hausdorff dimension of repellers of analytic expanding maps, using Bowen's pressure formula $P(-s\log|T'|) = 0$ where T is the map restricted to the repeller, and s is then the Hausdorff dimension of the repeller. This applies, for example, to hyperbolic Julia sets, and limit sets of convex cocompact Fuchsian groups. It offers an alternative to an earlier method of McMullen who used the transfer operator more directly to estimate solutions of $P(-s\log|T'|) = 0$, without using a zeta function.
- In (24), the same authors exhibit a super-exponentially converging numerical integration technique for analytic functions on the closed unit interval I. This relies on knowing the value of the function on a uniform set of points (the fixed points of T^n for $T(x) = 2x \pmod 1$). This can be generalised to functions on I^n. This method is efficient amongst algorithms which use values at evenly

distributed points, although other algorithms can be more efficient by choosing points to sample as part of the algorithm.

- Their paper (21) estimates the absolutely continuous invariant measure density function for a piecewise analytic expanding map T of the interval. The density function for the absolutely continuous invariant measure here is an analytic function, and it can be expanded as a Fourier series. They estimate the individual Fourier coefficients using periodic points of the transformation T.

- Pollicott and Rocha (45) calculate the determinant of the Laplacian on a manifold of constant negative curvature. The Laplacian is a second order self-adjoint linear operator which has countably many eigenvalues. The determinant of the Laplacian is a spectral invariant, formally defined via the meromorphic extension of a suitable complex function. The determinant is an important geometric characteristic. There are now other, more accurate, approximations of this quantity, see (60).

Transfer operators perform a kind of averaging on a function, where the points to sample are pre-images of an underlying transformation. They arose from statistical physics, and were first studied by Ruelle (47). His work established them as interesting objects in their own right. They were first applied to symbolic codings of dynamical systems, i.e. subshifts of finite type, on Banach spaces of Lipschitz continuous functions. This work was popularised by Bowen's book (7), which unified the ideas of Ruelle with Sinai's paper on Gibbs measures (56).

If a transfer operator has a spectral gap, then this leads to important results such as exponential decay of correlations, central limit theorems for distributions of time averages, and the existence of meromorphic zeta functions. In the case of transfer operators on subshifts of finite type, the Ruelle-Perron-Frobenius theorem (38), an analogue to the classical Perron-Frobenius theorem, in particular shows the transfer operator has a spectral gap. The key ingredient in the proof is what Ruelle called the 'basic inequality', and is now commonly referred to as the Lasota-Yorke inequality, however in different settings, it is much older (W. Doeblin and R. Fortet in 1937 showed quasi-compactness for a particular family of operators. In 1948 T. Ionescu and G. Marinescu generalised this result to obtain it for a larger family).

The Lasota-Yorke inequality is also used to show more abstract results, not just about transfer operators on subshifts of finite type, for example Nussbaum's formula for the essential spectral radius (36), and Hennion's theorem that in particular gives conditions for an operator to be quasi-compact. Other examples of its uses include the work of Hofbauer and Keller (18) who applied the inequality in the context of expanding interval maps with bounded variation functions, Pollicott (41) who used the inequality to study subshifts of finite type and Gibbs measures, and Liverani *et al* who considered operators for invertible transformations. The book of Hennion and Hervé (17) has a detailed description of the general method. Liverani's work provided a new perspective, where expanding maps were replaced by diffeomorphisms, by means of special distributional Banach spaces. However, there is not yet a theory of determinants for this setting.

An important application of transfer operators is to find absolutely continuous invariant measures for certain types of transformations. Such a measure exists, for example, when T is C^2 and expanding, in this case originally due to the work of Lasota and Yorke (28) which generalised earlier results for more specific maps. In these circumstances we can define a transfer operator $\mathcal{L}$ which has a positive fixed point ϕ_0, and whose dual preserves the Lebesgue measure. It then follows that $\int f \circ T \cdot \phi_0 dm = \int f \phi_0 dm$ where m is the Lebesgue measure, i.e. the measure $\phi_0 dm$ is T-invariant. In general such measures are called SRB measures after Sinai, Ruelle, and Bowen. See for example the books of Bowen (7), Parry and Pollicott (38), or Baladi (2).

An advantage of studying transfer operators is that they often preserve a space of higher regularity, depending on the underlying transformation T. Assume T is expanding, and the weight is $1/|T'|$. Under these assumption, for example, if T is C^k, $k \in \mathbb{N}$, then $\mathcal{L}(C^{k-1}) \subset C^{k-1}$. If T is $C^{1+\alpha}$ then $\mathcal{L}$ preserves C^α (α-Hölder continuous functions). The spectrum on the smaller spaces is often smaller, and eigenvalues in the smaller space are also in the larger one.

It is possible to define transfer operators in a variety of settings, but we are interested in the case when they work on Banach spaces of holomorphic functions, so that we can use the work of Grothendieck and also obtain convenient trace formulae.

At the beginning of the 20th century, Fredholm studied integral equations, where the general problem is to find a function f, given a function g and a kernel K (in some spaces) satisfying:

$$g(t) = \int K(s,t)f(s)ds.$$

Fredholm showed that the integral operator $Lf(t) = \int K(s,t)f(s)ds$ is trace-class, and defined

$$\det(I - zL) = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \operatorname{tr} L^n\right),$$

and this definition is natural because for finite matrices, it is an identity. For suitable conditions on the kernel, this gives us an entire function whose zeros are related to eigenvalues of the operator, and a product formula holds,

$$\det(I - zL) = \prod_{n=1}^{\infty}(1 - z\lambda_n),$$

where $(\lambda_n)_{n\in\mathbb{N}}$ is the eigenvalue sequence, in decreasing order of magnitude, repeated according to multiplicity. After the work of Riesz, it was known that the integral operator is compact, so has countably many eigenvalues, which arranged in a sequence, tend to zero. Schur had shown that the eigenvalue sequence for a continuous kernel is square-summable. It seemed that the smoother the kernel, the faster the rate of convergence. It became a general problem to find conditions on operators which ensures that their eigenvalue sequence is well behaved, e.g. belongs to ℓ^p for $0 < p < \infty$.

Grothendieck (13) defined the class of nuclear operators, and proved that any nuclear operator has a square-summable eigenvalue sequence. This was notable in that it applies in the more general setting of Banach spaces rather than Hilbert spaces. A linear operator $L : \mathcal{B}_1 \to \mathcal{B}_2$ between Banach spaces $\mathcal{B}_1$ and $\mathcal{B}_2$ is nuclear of order $p \geq 0$ if there exists sequences of elements $(u_n)_{n\in\mathbb{N}}$ in $\mathcal{B}_2$ of unit length, and linear functionals $(\ell_n)_{n\in\mathbb{N}}$ in $\mathcal{B}_1^\star$ such that $Lx = \sum_{n=1}^{\infty} \ell_n(x)u_n$ for all $x \in \mathcal{B}_1$, and $\sum_{n=1}^{\infty} \|\ell_n\|^q < \infty$ for all $q > p$. Nuclear operators of order $p < 2/3$ are trace class, so there is a bounded linear functional on nuclear operators $\mathcal{B} \to \mathcal{B}$ which generalises the usual finite dimensional trace functional, and allows us to define the Fredholm determinant as above.

It is also interesting that a formal link between transfer operators and integral operators exists because a transfer operator can be thought of as an integral operator with a kernel given by $K(x, y) = \delta(Ty - x)$ where δ is the Dirac Delta function.

Ruelle (47) first studied transfer operators on spaces of holomorphic functions, since this setting could employ Grothendieck's results. His paper establishes a general result about the transfer operator of the complexification of real analytic expanding maps on compact connected real analytic manifolds (that they are nuclear and have a particular equation for their traces), then applies this to express the Artin-Mazur and Smale zeta-functions as Fredholm determinants of specific transfer operators. In this paper, Ruelle claimed that $\det(I - zL)$ for the transfer operator L has a power series whose coefficients obey $O(e^{-Cn^2})$ for some $C > 0$, however this replicated an earlier error of Grothendieck which Fried corrected to $O(e^{-Cn^{1+1/d}})$, where d comes from the set $U \subset \mathbb{C}^d$ the space of functions lives on. This is the key result enabling a variety of rapidly converging numerical approximations.

Zeta functions are meromorphic functions analogous to the famous Riemann zeta function in number theory - they are defined in terms of some objects we are interested in the cardinality or distribution of, and are useful for coding this information in the location of their zeros or poles. In dynamics, particular zeta functions can often be related to the Fredholm determinant of a particular transfer operator.

The use of zeta functions in dynamics arose from counting periodic orbits. For a map $f : X \to X$, we wish to consider the asymptotic behaviour of the numbers $\#\mathrm{Fix}(f^n)$, where $\mathrm{Fix}(f^n) = \{x \in X : f^n x = x\}$. The Artin-Mazur zeta function is defined by

$$\zeta(z) = \exp\left(\sum_{n=1}^{\infty} \frac{z^n}{n} \#\mathrm{Fix}(f^n)\right).$$

For example, for a subshift of finite type $\sigma : \Sigma_A \to \Sigma_A$, given by the matrix A, the Bowen-Lanford result tells us that $\zeta(z) = (\det(I - zA))^{-1}$, i.e. the zeta function is rational. This has a pole for $z = 1/\lambda$ where λ is the largest eigenvalue of A, so we can read off the entropy from this zeta function, which tells us the asymptotic growth rate of the number of period orbits.

The Artin-Mazur zeta function can be generalised by introducing a weighting function. If $f : M \to M$ is a map, and $\phi : M \to \mathbb{C}$ a weight function, then we can define the Ruelle zeta function by

$$Z(\phi) = \exp\left(\sum_{n=1}^{\infty} \frac{1}{n} \sum_{x \in \mathrm{Fix}(f^n)} \prod_{k=0}^{n-1} \phi(f^k x)\right)$$

Then this reduces to the Artin-Mazur zeta function for the constant $\phi(z) = z$.

Developments in geometry also prompted development of the use of zeta functions in dynamics. Selberg in 1956 defined a zeta function for geodesic flows on surfaces of constant negative curvature,

$$Z(s) = \prod_{\gamma} \prod_{k=0}^{\infty} \left(1 - e^{-(s+k)\ell(\gamma)}\right),$$

where the outer product is taken over all closed geodesics γ, and $\ell(\gamma)$ is the length of γ. This is an entire function with trivial zeros at $0, -1, -2, \ldots$ and zeros on $(0,1)$ and $\mathfrak{Re}\, s = 1/2$ which relate to the eigenvalues of the Laplacian operator on the surface. Smale generalised this zeta function to geodesic flows on surfaces of curvature -1, and proposed extending this to Axiom A flows.

In 1972 Bowen showed that the closed orbits of Axiom A flows are uniformly distributed in the non-wandering set with respect to the measure of maximum entropy. Parry (37) gave an alternative proof using zeta functions. For Axiom A flows, a partial meromorphic extension of the zeta function follows from the work of Parry and Pollicott. For special Anosov flows, where the stable and unstable foliations are real analytic, the work of Ruelle related these to determinants.

The Grothendieck theory tells us that a trace functional exists for nuclear operators, however many uses of this theory require us to have an expression for the trace of a transfer operator. Ruelle (47) gave a formula for the trace, and this has since been generalised to operators on wider classes of function spaces by Mayer (31), and Bandtlow and Jenkinson (5), who also prove that eigenvalue sequence for certain transfer operators is in some sense universal; the eigenvalue sequence is determined by the weights and inverse branches, not the particular 'favourable space' of functions the transfer operator acts on.

An outline of the remainder of this work is as follows:

- Chapter 2 contains standard definitions and results required in the remainder of the work.

- The subject of chapter 3 is Lyapunov exponents of random matrix products of families of matrices. Motivated by the work of Pollicott on approximating the Lyapunov exponent for discrete families of matrices, we investigate many of the details, and attempt a generalisation for the case of continuous families. The main result in this chapter is a new expression for the Lyapunov exponent of the continuous case, albeit of limited numerical use as it stands. We also arrive at a slightly modified version of the approximation of Pollicott for the discrete case.

- Chapter 4 is concerned with extracting information from the spectrum of a particular transfer operator about eigenvalues and eigenfunctions of the Laplacian on a surface of constant negative curvature. We know that the zeros of the Selberg zeta function determine the spectrum of the Laplacian. Moreover the Selberg zeta function can be realized as the determinant of a suitable transfer operator, thus allowing the possibility of accurate computations of such values. Moreover we are interested in the more ambitious programme of estimating the associated eigenvectors.

- The subject of chapter 5 is Mahler measures, their numerical approximations, and a couple of applications. Mahler measures are a special type of integral associated to polynomials, and in some cases we can apply the work of Jenkinson and Pollicott to enable numerical approximations. Their work uses dynamical determinants to derive a rapidly converging sequence converging to the value of an integral with C^ω integrand. If the zeros of the polynomial are unfortunately located, we might hope to split the integral up at the location of this zero, and divide by a polynomial which removes the zero. This hopefully only changes the answer by an amount given by a standard integral.

- In chapter 6, the final chapter, we examine entropy rates of hidden markov processes. A result of Han and Marcus relates this quantity to the action of certain maps on a simplex. We transfer this to a setting where we can use the

Grothendieck theory, give a new expression for the entropy rate in terms of the spectrum of a transfer operator, and finally give an algorithm for rapidly approximating the entropy rate.

Chapter 2

Background Material

Throughout we denote an open ball in a metric space (X, d) centred at $x \in X$ of radius $r > 0$ by $B_d(x, r)$, and when the metric used is unambiguous we write $B(x, r)$. Similarly for a normed vector space $(X, \|\cdot\|)$ we use the notation $B_{\|\cdot\|}(x, r)$.

2.1 Functional Analysis

Let $(\mathcal{B}, \|\cdot\|)$ denote a Banach space, and let $L : \mathcal{B} \to \mathcal{B}$ be a linear operator. The quantity

$$\|L\| = \sup\{\|Lx\| : x \in \mathcal{B}, \|x\| = 1\}$$

is called the *operator norm* of L. If this is finite, we say L is bounded, and we denote the set of all such bounded linear operators on $\mathcal{B}$ as $\mathcal{L}(\mathcal{B})$.

The *resolvent set* $\mathrm{res}(L)$ of L is the set of $z \in \mathbb{C}$ such that $L - zI$ has an inverse and $(L - zI)^{-1} \in \mathcal{L}(\mathcal{B})$. The *spectrum* of L, $\mathrm{sp}(L)$, is the complement of the resolvent,

$$\mathrm{sp}(L) = \{z \in \mathbb{C} : z \notin \mathrm{res}(L)\},$$

and is a compact subset of $\mathbb{C}$. The *eigenvalues* of L are defined as the points $\lambda \in \mathrm{sp}(L)$ where $L - \lambda I$ fails to be injective. The *multiplicity* of an eigenvalue is the dimension of the eigenspace $\{v \in \mathcal{B} : (L - \lambda I)v = 0\}$, if this equals 1 we say the eigenvalue is *simple*. An eigenvalue λ is isolated if there exists $\delta > 0$ such that $\mathrm{sp}(L) \cap B(\lambda, \delta) = \{\lambda\}$. Any isolated point of the spectrum is an eigenvalue.

The *spectral radius* $R(L)$ of L is

$$R(L) = \sup\{|z| : z \in \mathrm{sp}(L)\},$$

and the *essential spectral radius* $R_{\mathrm{ess}}(L)$ is the smallest number $R_{\mathrm{ess}} \geq 0$ such that any eigenvalue $\lambda \in \mathrm{sp}(L)$ with $|\lambda| > R_{\mathrm{ess}}$ is an isolated eigenvalue of finite multiplicity. If $R_{\mathrm{ess}}(L) < R(L)$, the operator is called *quasicompact*.

An operator $L \in \mathcal{L}(\mathcal{B})$ is *compact* if the image $\{Lx_n : n \in \mathbb{N}\}$ of any bounded sequence $(x_n)_{n \in N}$ in $\mathcal{B}$ contains a Cauchy sequence. An operator L is finite rank if the image $L(\mathcal{B})$ is finite-dimensional. Finite rank operators are compact.

The dual of $\mathcal{B}$ is the set $\mathcal{B}^\star$ of bounded linear functionals $\phi : \mathcal{B} \to \mathbb{C}$ and is a Banach space with the operator norm. The dual $L^\star : \mathcal{B}^\star \to \mathcal{B}^\star$ of an operator $L \in \mathcal{L}(\mathcal{B})$ is defined by

$$(L^\star \phi)(x) = \phi(Lx),$$

for all $\phi \in \mathcal{B}^\star$ and $x \in \mathcal{B}$.

We now define what it means for a map from an open subset of $\mathbb{C}$ to a Banach space to be holomorphic, which we require for using the perturbation theorem.

Let $U \subset \mathbb{C}$ be open and $\mathcal{B}$ a complex Banach space. Consider $f : U \to \mathcal{B}$. Fix $x \in U$. If there exists a bounded complex-linear mapping $L : \mathbb{C} \to \mathcal{B}$ such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - L(h)\|}{\|h\|} = 0$$

then we say L is the *Fréchet derivative* of f at x and denote it by $L = Df(x)$, and if this exists at every point $x \in U$ then we say f is *strongly holomorphic*.

If $f : U \to \mathcal{B}$ is locally bounded and the mapping

$$\lambda \mapsto \phi(f(x + \lambda))$$

is holomorphic in the usual sense at $\lambda = 0$ for all $x \in U$, and bounded linear functionals $\phi \in \mathcal{B}^\star$, we say f is *weakly holomorphic*.

Fortunately these two definitions are equivalent (for a proof, see for example the lectures notes of Omri Sarig (52)), so a function that satisfies either is called holomorphic. Holomorphic functions are closed under addition, multiplication by scalar, and composition. We can also extend this definition to say that a map $f : \mathcal{B}_1 \to \mathcal{B}_2$ between two Banach spaces is holomorphic if $f \circ g$ is holomorphic whenever $g : \mathbb{C} \to \mathcal{B}_1$ is holomorphic.

If we restrict ourselves to maps from $\mathbb{C}$ to a Banach space, we have the following result, which generalises some results in complex analysis of one variable. Proofs for these results can also be found in (52).

Proposition 2.1. *Let $U \subset \mathbb{C}$ be an open simply connected set, and $\mathcal{B}$ a Banach space. Then*

1. *(Morera's Theorem) $f : U \to \mathcal{B}$ is holomorphic if and only if for every closed smooth curve γ in U, we have*

$$\oint_\gamma f(z)dz = 0.$$

2. *(Cauchy's Integral Formula) If $f : U \to \mathcal{B}$ is holomorphic, any point $\xi \in U$, any closed smooth curve γ in U, and an integer $n \geq 0$, we have*

$$f^{(n)}(\xi) = \frac{n!}{2\pi i} \oint_\gamma \frac{f(z)}{(z - \xi)^{n+1}} dz.$$

3. *Consider a ball $B(z_0, r)$, for $r > 0$ and $z_0 \in \mathbb{C}$. A function $f : B(z_0, r) \to \mathcal{B}$*

is holomorphic if and only if there is a power series expansion

$$f(z) = f(z_0) + \sum_{n=1}^{\infty} a_n (z - z_0)^n, \quad |z - z_0| < r,$$

where $a_n \in \mathcal{B}$ *for all* n *and the series converges uniformly in the norm of* $\mathcal{B}$ *on compact subsets* $B(z_0, r)$.

Let $\{L_z\}_{z \in U}$ denote a family of operators in $\mathcal{L}(\mathcal{B})$, parameterised by $z \in U$ where $U \subset \mathbb{C}$ is an open neighbourhood of zero. The family is called an *analytic family* if the map $U \to \mathcal{L}(\mathcal{B})$ given by $z \mapsto L_z$ is holomorphic. An important result about analytic families is the following, for a proof see Katō's book (26), or again (52).

Theorem 2.2. *(Perturbation Theorem) Let* $\{L_z\}_{z \in U}$ *be an analytic family of operators, such that* L_0 *has an isolated simple eigenvalue* λ. *Then there exists an* $\epsilon > 0$, *and a family* $\{\lambda_z\}_{z \in B(0,\epsilon)} \subset \mathbb{C}$, *such that* $\lambda_0 = \lambda$, *the map* $z \mapsto \lambda_z$ *is holomorphic, and* λ_z *is an isolated simple eigenvalue of* L_z.

We also need to define a Banach space of holomorphic functions, which provides us with a space on which to define trace-class transfer operators. Let D be a open, bounded and connected set in $\mathbb{C}^n$.

Definition 2.3. We denote by $\mathrm{Hol}(D)$ the Fréchet space of holomorphic functions $D \to \mathbb{C}$, and by $C(\overline{D})$ the Banach space of continuous functions $\overline{D} \to \mathbb{C}$ with the supremum norm. We finally define the set

$$F(D) = \{f \in C(\overline{D}) \,:\, f|_D \in \mathrm{Hol}(D)\},$$

and the norm $\|\cdot\|_{F(D)}$ by

$$\|f\|_{F(D)} = \sup_{z \in \overline{D}} |f(z)|$$

for $f \in F(D)$.

Here, 'holomorphic' means holomorphic in the single variable case in each variable separately. Many of results in single variable complex analysis generalise, in particular Montel's theorem (see the book of Krantz (27)), which makes $(F(D), \|\cdot\|_{F(D)})$ into a Banach space. We will make use of Morera's theorem on each variable separately to show membership of this space.

Another common space is $L^2(D) \cap \mathrm{Hol}(D)$ which we denote $L^2_\omega(D)$. We have that $F(D) \subset L^2_\omega(D)$ but they do not co-incide because for example when $D = \mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ the function $f(z) = (z-1)^{-1/2}$ is not bounded in $\overline{\mathbb{D}}$ so is not a member of $F(\mathbb{D})$. However, by using the substitution $z = 1 + re^{i\theta}$, we have

$$\int_{\mathbb{D}} |f(z)|^2 dxdy \leq \int_0^2 \int_{-\pi}^{\pi} (1/r) r dr d\theta = 4\pi < \infty,$$

hence $f \in L^2_\omega(\mathbb{D})$.

2.2 Nuclear Operators

The theory in this section is mostly due to Grothendieck, see (13), and generalises the situation of integral operators considered by Fredholm. A relevant survey is in Ruelle's book (49), and many of the details may be found in the books of Ryan (51), Pietsche (39), and Schaefer (53).

As part of a very brief account of nuclear operators, we define tensor products of Banach spaces. One way to construct the tensor product of two Banach spaces X and Y is as follows. The tensor product $x \otimes y$ of elements $x \in X$ and $y \in Y$ is a linear functional on $B(X \times Y)$, the set of bilinear forms. It is given by $(x \otimes y)(A) = A(x, y)$. The tensor product $X \otimes Y$ is then the the subspace of $B(X \times Y)^\star$ spanned by the elements $x \otimes y$. An element $u \in X \otimes Y$ is given by $u = \sum_{i=1}^n x_i \otimes y_i$. This representation is not unique. We define the projective norm π on $X \otimes Y$ by

$$\pi(u) = \inf \left\{ \sum_{i=1}^{n} \|x_i\| \|y_i\| \, : \, u = \sum_{i=1}^{n} x_i \otimes y_i \right\},$$

and define the projective tensor product Banach space $X \hat{\otimes}_\pi Y$ to be the completion of $X \otimes Y$ with respect to the projective norm. We have the following useful formula for the projective norm on $X \hat{\otimes}_\pi Y$:

$$\pi(u) = \inf \left\{ \sum_{n=1}^{\infty} |\lambda_n| \, : \, u = \sum_{n=1}^{\infty} \lambda_n x_n \otimes y_n, \|x_n\| = 1, \|y_n\| = 1 \right\}.$$

There is a canonical operator $J : X^\star \hat{\otimes}_\pi Y \to \mathcal{L}(X, Y)$ of unit norm that associates $u = \sum_{n=1}^{\infty} \phi_n \otimes y_n$ with the operator $L_u : X \to Y$ given by $L_u(x) = \sum_{n=1}^{\infty} \phi_n(x) y_n$. We call the range of J the nuclear operators. They form a Banach space denoted by $\mathcal{N}(X, Y)$, with projective norm carried across. All nuclear operators are compact.

An important property of nuclear operators is that we can define their traces. Tensor product theory shows that there exists a functional $\mathrm{Tr} : \mathcal{N}(X) \to \mathbb{C}$, which is continuous and coincides with the usual trace for finite dimensional spaces. It is defined by

$$\mathrm{Tr}\, u = \sum_{n=1}^{\infty} \phi_n(x_n),$$

and this does not depend on the representation of u. Nuclear operators also form an ideal. If $S : X \to Y$ is a nuclear operator and $T \in \mathcal{L}(W, X)$ and $R \in \mathcal{L}(Y, Z)$, then RST is a nuclear operator.

We now turn to determinants of nuclear operators. For $u \in X^\star \hat{\otimes}_\pi X$, defining

$$\mathrm{Det}(1 + u) = 1 + \sum_{n=1}^{\infty} \sum_{i_1 < \cdots < i_n} \det(\phi_{i_l}(x_{i_k}))_{l,k=1}^{n} \tag{2.1}$$

makes sense because there is a similar formula in the finite case.

Proposition 2.4. *We have the following facts about the determinant and trace:*

1. *$u \mapsto \mathrm{Det}(1+u)$ is an entire holomorphic function.*

2. *$\mathrm{Det}(1+u) \neq 0$ if and only if L_u is invertible in $\mathcal{L}(X)$.*

3. *If λ is an eigenvalue of multiplicity n for L_u then λ^{-1} is a zero of order n of $z \mapsto \mathrm{Det}(1-zu)$.*

4. *The identity*
$$\mathrm{Det}(1-zu) = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{Tr}\, u^n\right) \tag{2.2}$$
holds where the power series converges for a neighbourhood of zero.

The rate in which the terms of u tend to zero gives further information, so we state the following definition.

Definition 2.5. An operator $\mathcal{L} : X \to X$ on a complex Banach space X is called nuclear of order $p \geq 0$ (alternatively, p-nuclear) if there exists sequences $(\lambda_n)_{n\in\mathbb{N}} \in \mathbb{C}$, $(\phi_n)_{n\in\mathbb{N}} \in X^\star$ and $(x_n)_{n\in\mathbb{N}} \in X$ with $\|\phi_n\| = 1$ and $\|x_n\| = 1$ for all n such that for all $x \in X$:
$$\mathcal{L}x = \sum_{n=1}^{\infty} \lambda_n \phi_n(x) x_n,$$
and $\sum_{n=1}^{\infty} |\lambda_n|^q < \infty$ for all $q > p$. It is called strongly nuclear if it is nuclear of order zero.

For each $p \geq 0$, the p-nuclear operators form an ideal.

Proposition 2.6. *If u is p-nuclear for $p \leq 2/3$, then the eigenvalue sequence $(\lambda_n)_{n\in\mathbb{N}}$ of L_u has $\sum_{n=1}^{\infty} |\lambda_n|^p < \infty$ and*
$$\mathrm{Tr}\, u = \sum_{n=1}^{\infty} \lambda_n.$$
Furthermore we have the following usual connection between the determinant and the trace:
$$\mathrm{Det}(1-zu) = \prod_{n=1}^{\infty} (1 - z\lambda_n).$$

Example 2.7. Define the linear operator $L_s : F(\mathbb{D}) \to F(\mathbb{D})$ by $Lf(z) = f(z/s)$ for $s > 1$. Define the sequence $(u_n)_{n\in\mathbb{N}_0}$ of functions $\overline{\mathbb{D}} \to \overline{\mathbb{D}}$ by $u_n(z) = z^n$, which is a basis for $F(\mathbb{D})$, and the dual basis $(\phi_n)_{n\in\mathbb{N}_0}$ by $\phi_n(u_m) = 1$ if $n = m$, and 0 otherwise. Clearly $\|u_n\|_{F(\mathbb{D})} = \sup_{z\in\overline{\mathbb{D}}} |u_n(z)| = 1$ and $\|\phi_n\| = 1$ for all $n \in \mathbb{N}_0$.

Fix an arbitrary $f \in F(\mathbb{D})$, and write it as $f = \sum_{n=0}^{\infty} a_n u_n$ where $(a_n)_{n\in\mathbb{N}_0}$ is a sequence in $\mathbb{C}$. We can then write the operator as

$$L_s f(z) = \sum_{n=0}^{\infty} a_n u_n(z/s) = \sum_{n=0}^{\infty} a_n (1/s)^n u_n(z) = \sum_{n=0}^{\infty} (1/s)^n \phi_n(f) u_n(z).$$

Then L_s is of the form $L_s f(z) = \sum_{n=0}^{\infty} \lambda_n \phi_n(f) u_n(z)$ where ϕ_n and u_n are normalised, and $\lambda_n = 1/s^n$. Now $\sum_{n=0}^{\infty} |\lambda_n|^p = \sum_{n=0}^{\infty} (1/s^p)^n < \infty$ provided $1/s^p < 1$, i.e. $p > 0$. This shows L_s is strongly nuclear.

Furthermore, since $L_s(u_n) = 1/s^n u_n$, we see that $1/s^n$ is an eigenvalue for each $n \geq 0$. From this and the strong nuclearity we have that $\operatorname{Tr} L_s = \sum_{n=0}^{\infty} s^{-n} = 1/(1 - s^{-1})$. Since $L_s^n = L_{s^n}$, we can see that $\operatorname{Tr} L_s^n = \operatorname{Tr} L_{s^n} = 1/(1 - s^{-n})$.

The standard ways of showing an operator is p-nuclear or strongly nuclear is using Cauchy's theorem to obtain the linear functionals, or using the fact the space of nuclear operators is an ideal to factor an operator L into a composition RST where R and T are any bounded linear operators, and S is an operator already known to be nuclear.

We also have a source of nuclear operators because some spaces are nuclear, that is spaces where any bounded linear operator from that space to any Banach space is nuclear. An example of a nuclear space is the Fréchet space $\operatorname{Hol}(D)$ where $D \subset \mathbb{C}^d$ is a bounded, open, connected set.

We can give a power series expansion of the determinant using equation 2.2,

$$\operatorname{Det}(1 - zu) = 1 + \sum_{n=1}^{\infty} b_n z^n,$$

with the terms b_n worked out as follows.

$$\begin{aligned}
\operatorname{Det}(1-zu) &= 1+\sum_{n=1}^{\infty}\frac{1}{n!}\left(-\sum_{m=1}^{\infty}\frac{z^m}{m}\operatorname{Tr}u^m\right)^n \\
&= 1+\sum_{n=1}^{\infty}\frac{(-1)^n}{n!}\sum_{p=1}^{\infty}z^p\sum_{\substack{m_1,\ldots,m_n\in\mathbb{N}\\ m_1+\cdots m_n=p}}\prod_{i=1}^{n}\frac{\operatorname{Tr}u^{m_i}}{m_i} \\
&= 1+\sum_{n=1}^{\infty}z^n b_n,
\end{aligned}$$

where

$$b_n=\sum_{\substack{n_1,\ldots,n_m\in\mathbb{N}\\ n_1+\cdots n_m=n}}\frac{(-1)^m}{m!}\prod_{i=1}^{m}\frac{\operatorname{Tr}u^{n_i}}{n_i}. \tag{2.3}$$

This is central to using this theory to construct quickly converging numerical approximations to various quantities. This requires bounding the terms b_n (and also requires a suitable expression for the trace, discussed later). Assume u is nuclear and can be written in the form $u=\sum_{n=1}^{\infty}a_n\phi_n\otimes x_n$ with $\|\phi_n\|=\|x_n\|=1$ for all $n\in\mathbb{N}$, and the terms $a_n\in\mathbb{C}$ are controlled by

$$0<|a_n|<A\alpha^{n^{1/d}} \tag{2.4}$$

for constants $A>0$, $0<\alpha<1$ and $d\in\mathbb{N}$. First, equation 2.1 gives

$$b_n=\sum_{i_1<\cdots<i_n}a_{i_1}a_{i_2}\cdots a_{i_n}\det(\phi_{i_l}(x_{i_k}))_{l,k=1}^{n}. \tag{2.5}$$

By a result of Hadamard (14) we have that any $N\times N$ matrix with entries of modulus at most 1 has determinant with modulus bounded by $N^{N/2}$, hence

$$|b_n|\le n^{n/2}A^n\sum_{i_1<\cdots<i_n}\alpha^{{i_1}^{1/d}+{i_2}^{1/d}+\cdots+{i_n}^{1/d}}.$$

The sum in the right hand side is a coefficient of the power series of the function

$f_\alpha(z) = 1 + \sum_{n=1}^{\infty} \beta_n(\alpha) z^n = \prod_{i=1}^{\infty}(1 + \alpha^{i^{1/d}} z)$, so that

$$|b_n| \leq n^{n/2} A^n \beta_n(\alpha). \tag{2.6}$$

Fried (12) showed that $\beta_n(\alpha) = O(\delta^{n^{1+1/d}})$ where $0 < \delta < 1$, and Jenkinson and Pollicott in (23) reproduce Fried's analysis. Fried used Cauchy's estimate, then examined the number of zeroes in the ball $|z| < r$, to show that $\beta_n(\alpha) \leq r^{-n} \exp(a_\alpha P(\log r))$ for all $r > 0$ where $a_\alpha = (-\log \alpha)^{-d} > 0$, and $P(x) = \sum_{j=0}^{d+1} \frac{d!}{j!} x^j$. Define $r_n = \exp((n/a_\alpha)^{1/d})$. Then $\beta_n(\alpha) \leq r_n^{-n} \exp(a_\alpha P(\log r_n)) = \mathcal{O}(\epsilon^{n^{1+1/d}})$ for any $1 > \epsilon > \alpha^{d/(d+1)}$. Substituting this into equation 2.6 gives $b_n = \mathcal{O}(\delta^{n^{1+1/d}})$ for any $\epsilon < \delta < 1$. This is summarised in the following proposition.

Proposition 2.8. *If an operator L is nuclear, and can be written in the form $Lf = \sum_{n=1}^{\infty} a_n \phi_n(f) u_n$, with $\|\phi_n\| = \|u_n\| = 1$ and $0 < |a_n| < A\alpha^{n^{1/d}}$, $A > 0$, $0 < \alpha < 1$, then the determinant $\det(1 - zL) = 1 + \sum_{n=1}^{\infty} b_n z^n$ has coefficients which decrease super-exponentially fast. We have $b_n = \mathcal{O}(\delta^{n^{1+1/d}})$ for any $\alpha^{d/(d+1)} < \delta < 1$.*

Use of the Hadamard estimate doesn't affect the rate of the convergence but it suggests that we lose information which might be used to tighten up the constant implied in the $\mathcal{O}$ notation. For the case where $d = 1$, we can write $|b_n| \leq n^{n/2} A^n \alpha^{n^2} \alpha^{P(-n \log \alpha)}$ where $P(x) = 1 + x + x^2/2$.

The bound on the $(a_n)_{n\in\mathbb{N}}$ sequence is usually established in practice using Cauchy estimates. Alternatively, if a similar bound can be established on the eigenvalues of the operator instead, we can obtain a similar result about the rate of convergence of the determinant.

Proposition 2.9. *If a strongly nuclear operator $\mathcal{L}$ has eigenvalues $(\lambda_n)_{n\in\mathbb{N}}$, and we have a bound*

$$0 < |\lambda_n| < A\alpha^{n^{1/d}}$$

for all n, for some $A > 0$ and $0 < \alpha < 1$, then we have $b_n = \mathcal{O}(\delta^{n^{1+1/d}})$ for any

$\alpha^{d/(d+1)} < \delta < 1$.

Proof. Since the operator is strongly nuclear we have

$$1 + \sum_{n=1} b_n z^n = \det(1 - z\mathcal{L}) = \prod_{i=1}^{\infty}(1 - z\lambda_n) = 1 + \sum_{n=1}^{\infty}(-1)^n z^n \sum_{i_1<\cdots<i_n} \lambda_{i_1} \cdots \lambda_{i_n},$$

so $b_n = (-1)^n \sum_{i_1<\cdots<i_n} \lambda_{i_1} \cdots \lambda_{i_n}$. We can then bound

$$|b_n| \leq A^n \sum_{i_1<\cdots<i_n} \alpha^{{i_1}^{1/d}+\cdots {i_n}^{1/d}},$$

and perform Fried's analysis as before. □

Example 2.10. Returning to example 2.7, we can write

$$\det(1 - zL_s) = \exp\left(-\sum_{n=1}^{\infty} z^n \frac{1}{n(1 - s^{-n})}\right),$$

and also notice that the determinant of the matrix in equation 2.5 is now 1, hence the terms in the power series for the Fredholm determinant are in this case bounded by $|b_n| \leq s^{-n^2} s^{-P(n \log s)}$. Here L_s is a composition operator, and the underlying map is a contraction, and we observe that the stronger the contraction ratio (the larger we make s) then the faster the rate of convergence of the determinant.

2.3 Transfer Operators

Transfer operators are usually of the following flavour. Consider a transformation $T : X \to X$, on some space X, with $T^{-1}(x)$ countable or finite for each $x \in X$, and let $g : X \to \mathbb{C}$ be a *weight* function such that $\sum_{y \in T^{-1}(\{x\})} g(y)$ is convergent. Let $\mathcal{B}$ be a suitable Banach space of functions on $X \to \mathbb{C}$. Then the *transfer operator*

$\mathcal{L}_g : \mathcal{B} \to \mathcal{B}$ is defined as

$$\mathcal{L}_g f(x) = \sum_{y\,:\,Ty=x} f(y)g(y).$$

The archetypal setting for transfer operators is subshifts of finite type. The reader is referred to the book of Parry and Pollicott (38) for more information and proofs. This setting arises for example as a result of looking at the coding of an expanding map on a Markov partition, or lattice spin systems in statistical physics. Let A denote a $k \times k$ matrix with each entry either zero or one, which is furthermore aperiodic, i.e. some iterate A^n is strictly positive. Let $S = \{1, \ldots, k\}$ and

$$\Sigma_A^+ = \{x \in S^{\mathbb{N}} \,:\, A(x_n, x_{n+1}) = 1 \text{ for all } n\}$$

with the Tychonov product topology. This topology is generated by cylinders, let $j \geq 1$ and $x_1, \ldots, x_n \in S$, then the cylinder $Z(j; x_1, x_2, \ldots, x_n)$ is given by

$$Z(j; x_1, x_2, \ldots, x_n) = \{y \in \Sigma_A^+ \,:\, y_j = x_1, y_{j+1} = x_2, \ldots, y_{j+n} = x_n\}.$$

Note that a cylinder might be empty. The topology is also metrizable. Given $0 < \theta < 1$ define $d_\theta(x, y) = \theta^N$ where $N = \min\{k \,|\, x_k \neq y_k\}$. The shift $\sigma \,:\, \Sigma_A^+ \to \Sigma_A^+$ is then the transformation which deletes the first term in the sequence, $\sigma x = y$ where $y_n = x_{n+1}$. For a function $f \,:\, \Sigma_A^+ \to \mathbb{C}$ and $n \geq 0$ define

$$\mathrm{var}_n f = \sup\{|f(x) - f(y)| \,:\, x_i = y_i, 0 \leq i \leq n\}$$

and

$$|f|_\theta = \sup\{\theta^{-n} \mathrm{var}_n f \,:\, n \geq 0\}.$$

This is only a pseudo-norm, so define $\|f\|_\theta = |f|_\infty + |f|_\theta$. Let

$$\mathcal{F}_\theta^+ = \{f : \Sigma_A^+ \to \mathbb{C} : f \text{ continuous}, \ \|f\|_\theta < \infty\},$$

the space of d_θ-Lipschitz continuous functions.

Let $g \in \mathcal{F}_\theta^+$. The Ruelle transfer operator $\mathcal{L}_g : \mathcal{F}_\theta^+ \to \mathcal{F}_\theta^+$, a bounded linear operator, is defined by

$$\mathcal{L}_g f(x) = \sum_{\sigma y = x} g(y) f(y).$$

The following important theorem, a generalisation of the classical Perron-Frobenius theorem, is due to Ruelle (48). In other settings for transfer operators, the hope is that similar results hold.

Theorem 2.11. *(Ruelle-Perron-Frobenius) Let $g \in \mathcal{F}_\theta^+$ be real valued and strictly positive.*

1. *There is a simple maximal positive eigenvalue β of $\mathcal{L}_g$ with corresponding strictly positive eigenfunction $h \in \mathcal{F}_\theta^+$.*

2. *The remainder of the spectrum of $\mathcal{L}_g$ is contained in a disc of radius strictly less than β. Hence $\mathcal{L}_g$ is quasicompact.*

3. *There is a unique σ-invariant probability measure μ such that $\mu(\mathcal{L}_g(v)) = \beta\mu(v)$ for all $v \in C(\Sigma_A^+)$, the set of continuous functions on Σ_A^+.*

4. *$\beta^{-n}\mathcal{L}_g^n v \to h\,\mu(v)/\mu(h)$ uniformly for all $v \in C(\Sigma_A^+)$.*

If g is real and strictly positive, the measure μ from this theorem is an equilibrium state for $\log g$, and

$$\beta = e^{P(\log g)},$$

where $P(\cdot)$ is the pressure of a function (see the book of Walters (61) for a definition and important results).

Transfer operators have also been extensively studied on Banach spaces of holomorphic maps. Less smooth settings have been studied, but the theory of Grothendieck discussed in the next section ensures that the operators will be trace class. Here we have that the transfer operator is a sum of composition operators, multiplied by a weight. So we consider an open, bounded and connected domain $D \subset \mathbb{C}^d$, a finite or countable indexing set I, a family of functions (often called inverse branches) $\{T_i\}_{i\in I}$ where $T_i : D \to D$ is holomorphic, and another family of functions $\{w_i\}_{i\in I}$ with $w_i : D \to \mathbb{C}$ holomorphic. The transfer operator can be defined on a variety of Banach spaces of holomorphic functions, with the case that Ruelle studied being $F(D)$. We will also use this space as our setting. The transfer operator $\mathcal{L} : F(D) \to F(D)$ is defined, for $f \in F(D)$, $z \in D$ by

$$\mathcal{L}f(z) = \sum_{i\in I} f(T_i(z))w_i(z).$$

We also require the condition that

$$\overline{\bigcup_{i\in I} T_i(D)} \subset D, \tag{2.7}$$

in order to get unique fixed points of the inverse branches, and to make sure the operator is 'analyticity improving'. See figure 2.1.

Ruelle (47) proved that the transfer operator $\mathcal{L}$ is nuclear. Bandtlow and Jenkinson (5) proved nuclearity for a wide variety of Banach spaces, including $F(D)$. Mayer (33) looked at the case where inverse branches T_i are on an infinite dimensional Banach space, and gave technical conditions (including that the derivatives of the branches are nuclear) for the transfer operator to be nuclear.

Ruelle obtained (and similarly Bandtlow and Jenkinson generalised) a formula for the trace of the transfer operator. First a simpler situation must be considered. Consider holomorphic maps $T : D \to D$ and $w : D \to \mathbb{C}$, and the transfer operator

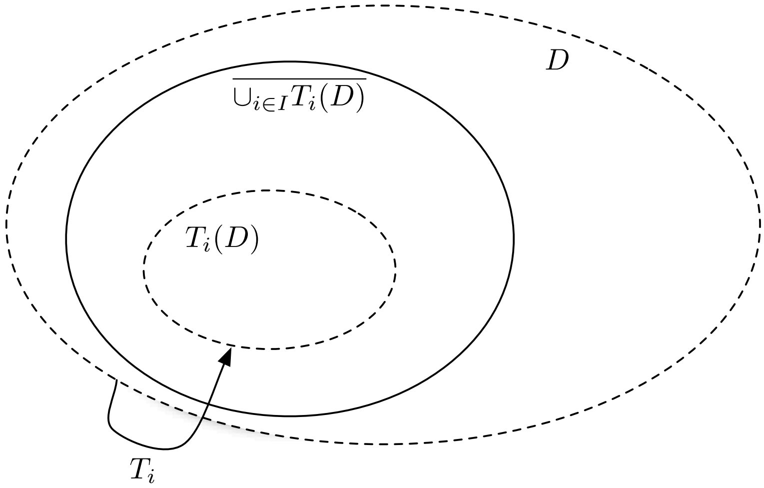


Figure 2.1: The condition $\overline{\cup_{i\in I} T_i(D)} \subset D$.

$\mathcal{M} : F(D) \to F(D)$ given by $\mathcal{M}f(z) = f(T(z))w(z)$ for $f \in F(D)$ and $z \in D$. We again require $\overline{T(D)} \subset D$. Let $z_0 \in D$ denote the unique fixed point $T(z_0) = z_0$. We have the following formula for the trace,

$$\operatorname{Tr} \mathcal{M} = \frac{w(z_0)}{\det(I - DT(z_0))} \tag{2.8}$$

The formula for $\operatorname{Tr} \mathcal{L}^n$ is built up using this simpler formula. Fix $\underline{i} = (i_1, \ldots, i_n) \in I^n$. Define $T_{\underline{i}} = T_{i_1} \circ T_{i_2} \circ \cdots \circ T_{i_n}$, let $z_{\underline{i}} \in D$ be its unique fixed point, and define the map $w_{\underline{i}} : D \to \mathbb{C}$ by $w_{\underline{i}}(z) = \prod_{m=1}^{n} w_{i_m}(T_{i_{m+1}} \cdots T_{i_n} z)$. Then the trace formula for $\mathcal{L}^n$ follows from equation 2.8,

$$\operatorname{Tr} \mathcal{L}^n = \sum_{\underline{i}\in I^n} \frac{w_{\underline{i}}(z_{\underline{i}})}{\det(I - DT_{\underline{i}}(z_{\underline{i}}))}. \tag{2.9}$$

Example 2.12. We give an example illustrating the use of the theory in the paper of Jenkinson and Pollicott (23). Let $U = B(1/2, 3/2)$, and $T_i : U \to U$, $i = 1, 2$ be given by $T_1(z) = z/3$ and $T_2(z) = z/3 + 2/3$. Define a transfer operator $\mathcal{L}_s : F(U) \to F(U)$ by $\mathcal{L}_s f(z) = 3^{-s}(f(T_1 z) + f(T_2 z))$. Notice that $\mathcal{L}_s 1 = 3^{-s}2$, so that 1 is an eigenvalue when $3^{-s}2 = 1$, i.e. $s = \log 2/\log 3$. Of course this is the dimension of the middle third Cantor set X, which is the attractor for the IFS given by the two maps T_1 and T_2, which are inverse branches of the dynamical system $T : X \to X$ given by $T(x) = 3x \,(\mathrm{mod}\, 1)$. For transfer operators of this type, there is a maximal real eigenvalue given by $e^{P(-s\log|T'|)}$, and if it is 1 then this must mean $P(-s\log|T'|) = 0$, and by Bowen's elegant formula, s is then the dimension of the repeller for T.

In general, X is a compact set in a C^∞ manifold which is the attractor for a dynamical system $T : X \to X$, such that T is expanding, analytic, conformal, and locally maximal, i.e. for a sufficiently small open neighbourhood U of X, we have $\cap_{n=1}^{\infty} T^{-n}U = X$. We also require that the dynamics of T are coded by a subshift of finite type. The point of these hypotheses is so that it is always possible to find inverse branches which behave nicely, and can be complexified, so that there is a rich setting to define transfer operators, including being able to use nuclear operator theory. As mentioned in the introduction, the determinant associated with the transfer operator provides the necessary information to find this eigenvalue, and provides a method for a numerical approximation.

Remark 2.13. If $\mathcal{L}_t : F(U) \to F(U)$ is a holomorphic family of transfer operators defined by $\mathcal{L}_t f(z) = \sum_i f(T_i z)\exp(tw(T_i z))$, and μ is a $\mathcal{L}_0$-invariant measure, then there is a nice connection between w and the spectrum of $\mathcal{L}_0$. Let $h_t \in F(U)$ be the family of eigenvectors corresponding to eigenvalues λ_t we get from applying the perturbation theorem with $\lambda_0 = h_0 = 1$. Differentiate $\mathcal{L}_t h_t = \lambda_t h_t$ with respect to t, to get $\mathcal{L}_t(h_t' + h_t w) = \lambda_t' h_t + \lambda_t h_t'$. At $t = 0$, this reduces to $\mathcal{L}_0(h_t'|_{t=0} + w) = \lambda_t'|_{t=0} + h_t'|_{t=0}$. Applying μ, using the $\mathcal{L}_0$ invariance, and simplifying gives the equation

$$\int w\, d\mu = \lambda_t'|_{t=0}.$$

This connection is used, for example, by Jenkinson and Pollicott (24) to construct an algorithm for working out integrals of analytic functions with respect to the

Lebesgue measure.

Remark 2.14. Sometimes the weight function is of the form $|w|$ where w is a holomorphic function. The imaginary part of $|w|$ is always zero, so it cannot satisfy the Cauchy-Riemann equations, so is not holomorphic. However, a trick can be performed, which allows us to use the Grothendieck machinery again. This doesn't quite come for free because it increases the dimension of the domain we work in, so the convergence of the power series of the determinant isn't as fast.

Let $T : \mathbb{C} \to \mathbb{C}$ be the underlying transformation. Write

$$T(x+iy) = T_1(x,y) + iT_2(x,y),$$

for $x, y \in \mathbb{R}$, and complexify each of $T_1 : \mathbb{R}^2 \to \mathbb{R}$ and $T_2 : \mathbb{R}^2 \to \mathbb{R}$ to get a function $\tilde{T} : \mathbb{C}^2 \to \mathbb{C}^2$ where

$$\begin{aligned} \tilde{T}(x_1+ix_2, y_1+iy_2) \;=\; & (T_{11}(x_1+ix_2, y_1+iy_2) + iT_{12}(x_1+ix_2, y_1+iy_2), \\ & T_{21}(x_1+ix_2, y_1+iy_2) + iT_{22}(x_1+ix_2, y_1+iy_2)). \end{aligned}$$

We have

$$\begin{aligned} T_{11}(x_1+0i, y_1+0i) &= T_1(x_1, y_1) \\ T_{12}(x_1+0i, y_1+0i) &= 0 \\ T_{21}(x_1+0i, y_1+0i) &= T_2(x_1, y_1) \\ T_{22}(x_1+0i, y_1+0i) &= 0, \end{aligned}$$

for all $x_1, y_1 \in \mathbb{R}$. These relations tell us derivatives such as $\partial T_{22}/\partial x_1 = 0$, $\partial T_{11}/\partial x_1 = \partial T^r/\partial x_1$. Using the fact that T_1 and T_2 satisfy the Cauchy-Riemann equations in each variable separately, we may consider the value of the matrix

$$D\tilde{T}(x_1 + 0i, y_1 + 0i) = \begin{pmatrix} \frac{\partial \tilde{T}_1}{\partial x}(x_1, y_1) & \frac{\partial \tilde{T}_1}{\partial y}(x_1, y_1) \\ \frac{\partial \tilde{T}_2}{\partial x}(x_1, y_1) & \frac{\partial \tilde{T}_2}{\partial y}(x_1, y_1) \end{pmatrix}$$

and using the correspondence between complex numbers and real 2x2 matrices, we get

$$D\tilde{T}(x_1, y_1) = \begin{pmatrix} \frac{\partial T_1}{\partial x}(x_1, y_1) & \frac{\partial T_1}{\partial y}(x_1, y_1) \\ \frac{\partial T_2}{\partial x}(x_1, y_1) & \frac{\partial T_2}{\partial y}(x_1, y_1) \end{pmatrix},$$

hence

$$\det(I - D\tilde{T}(x_1, y_1)) = |1 - T'(x_1 + iy_1)|^2.$$

We may verify that the map $\tilde{T}$ satisfies the Cauchy-Riemann equations in each variable separately, and also that the fixed points of T correspond bijectively to the fixed points of $\tilde{T}$ via the natural embedding of $\mathbb{C}$ in $\mathbb{C}^2$. The value of the complexified weight function at these fixed points will be remain the same, so we may thus simplify the trace formula.

Chapter 3

Lyapunov Exponents of Random Matrix Products

3.1 Introduction

Let I be a compact topological space equipped with a probability measure ρ. Consider a family of strictly positive non-singular $d \times d$ matrices $\{A_\alpha\}_{\alpha \in I}$, such that each component varies continuously with α.

The Lyapunov exponent λ of this system is defined to be

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \int_I \int_I \cdots \int_I \log \|A_{\alpha_1} A_{\alpha_2} \cdots A_{\alpha_n}\| d\rho(\alpha_1) d\rho(\alpha_2) \cdots d\rho(\alpha_n). \qquad (3.1)$$

Because the sequence $(\lambda_n)_{n \in \mathbb{N}}$ given by

$$\lambda_n = \int_I \cdots \int_I \log \|A_{\alpha_1} \cdots A_{\alpha_n}\| d\rho(\alpha_1) \cdots d\rho(\alpha_n)$$

is easily seen to be sub-additive ($\lambda_{n+m} \leq \lambda_n + \lambda_m$ for all $n, m \in \mathbb{N}$), we may apply the well-known lemma (see for example the book of Walters (61)), to give that $\lambda \in [-\infty, \infty)$. To exclude the possibility that $\lambda = -\infty$, we use that there exists constants $B > 0$, $C \in \mathbb{R}$ such that

$$\|A_{\alpha_1} \cdots A_{\alpha_n}\| \geq Be^{Cn}$$

for all $n \in \mathbb{N}$ and all $\alpha \in I^n$, which immediately gives $\lambda \geq C$. To show the inequality, if the smallest entry of a $d \times d$ matrix A is a, then $\|Ax\| \geq \|x\|_1 a d^{1/2}$ for all $x \in \mathbb{R}^d_+$, where $\mathbb{R}^d_+$ denotes elements of $\mathbb{R}^d$ with strictly positive components. Let a_α be the smallest entry for A_α, and $a = \inf_{\alpha \in I} a_\alpha$. Each entry in any product $A_{\alpha_1} \cdots A_{\alpha_n}$ is then at least $d^{n-1}a^n$. Hence, using that $\|A\| = \sup\{\|Ax\| \, : \, \|x\| = 1, x \in \mathbb{R}^d_+\}$ for strictly positive A, we have

$$\|A_{\alpha_1} \cdots A_{\alpha_n}\| \geq \sup_{\|x\|=1} \|x\|_1 d^{-1/2} d^n a^n = Be^{Cn},$$

for some constants B, C. Showing an upper bound is simpler, since

$$\lambda_n \leq n \int_I \log \|A_\alpha\| d\rho(\alpha),$$

for all n, we have $\lambda \leq \int_I \log \|A_\alpha\| d\rho(\alpha)$.

Consider the case where $I = \{1, \ldots, m\}$ is equipped with the discrete topology, and ρ is a probability measure on I, so is given by $\rho = \sum_{i=1}^m p_i \delta_i$ where each $p_i \geq 0$, $p_1 + \cdots + p_n = 1$, and δ_i is the Dirac delta function at i, i.e. $\delta_i(A) = \chi_A(i)$. Then equation 3.1 becomes

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \sum_{\underline{i} \in I^n} p_{i_1} p_{i_2} \cdots p_{i_n} \log \|A_{i_1} A_{i_2} \cdots A_{i_n}\|,$$

which matches the discrete case studied by Pollicott (43), and motivates the work in this chapter, in which we hope to follow the ideas of Pollicott but in the more general case. The paper of Pollicott (43) explains the interest of the Lyapunov exponent of the discrete case, however it is natural to examine the continuous case.

We would like to recast the quantity λ as something involving a suitably defined transfer operator, to find a new expression for the Lyapunov exponent. We can then invoke the tools of Grothendieck and write the Lyapunov exponent in terms of the determinant, involving periodic points of the transformations used to define the transfer operator.

Transfer operators sum or integrate over a family of composition operators. The tools for calculating the trace of these transfer operators rely heavily on complex analysis. We therefore need to study a family of maps which encode spectral information about the matrices, and are nice from a complex analysis point of view. These maps are the multi-dimensional linear fractional transformations, which are used in the study of multi-dimensional continued fractions.

First note that if we set

$$A'_\alpha = a_\alpha A_\alpha$$

where a_α is some scaling variable which depends continuously on α then we see that the Lyapunov exponent λ' of this family is

$$\begin{aligned}\lambda' &= \lim_{n\to\infty}\frac{1}{n}\int_I\int_I\cdots\int_I \log\|A_{\alpha_1}A_{\alpha_2}\cdots A_{\alpha_n}\| \\ &\quad + \sum_{i=1}^{n}\log|a_{\alpha_i}|d\rho(\alpha_1)d\rho(\alpha_2)\cdots d\rho(\alpha_n) \\ &= \lambda + \int_I \log|a_\alpha|d\rho(\alpha),\end{aligned}$$

so if we set $a_\alpha = |\det A_\alpha|^{-1/d}$ then $\det A'_\alpha = 1$ or -1 for all α. Therefore, without loss of generality, assume all the matrices in the family have determinant 1. We can recover the original Lyapunov exponent from this assumption using the above equation.

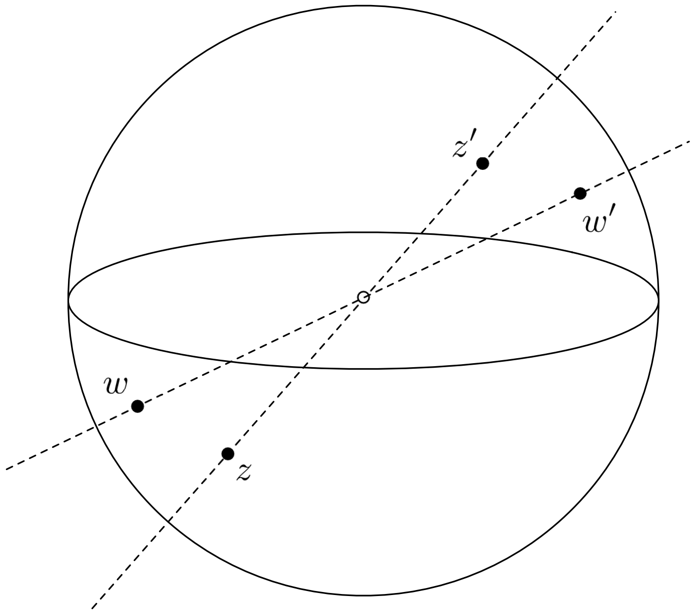


Figure 3.1: $z \sim z'$ and $w \sim w'$ because the lines which connect them pass through the origin.

3.2 A Family of Maps

We first examine the manifold $\mathbb{C}P^{d-1}$, the complex projective space, and show that it naturally gives rise to the linear fractional transformation associated with a matrix A_α.

For $0 \neq z, w \in \mathbb{C}^d$, we say that $z \sim w$ if there exists $\lambda \in \mathbb{C}$ such that $z = \lambda w$, see figure 3.1. We write $[z] = \{w \in \mathbb{C}^d : w \sim z\}$, and define $\mathbb{C}P^{d-1}$ as the quotient space

$$\mathbb{C}P^{d-1} = (\mathbb{C}^d - \{0\})/\sim = \{[z] : z \in \mathbb{C}^d - \{0\}\}.$$

We equip it with the quotient topology, that is defining $\pi : \mathbb{C}^d \to \mathbb{C}P^{d-1}$ by $\pi z = [z]$, we declare πU to be open in $\mathbb{C}P^{d-1}$ whenever $U \subset \mathbb{C}^d$ is open. For readability $[z_1, \ldots, z_d]$ is understood to mean $[(z_1, \ldots, z_d)]$.

We put an atlas on this space to make it a complex manifold. Define subsets

$U_i = \{[z_1, \dots, z_d] : z_i \neq 0\}$ and then maps $\phi_i : U_i \to \mathbb{C}^{d-1} - \{0\}$ by

$$\phi_i[z_1, \dots, z_d] = z_i^{-1}(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_d).$$

The sets U_i cover $\mathbb{C}P^{d-1}$, and that the maps ϕ_i are well defined bijections. We show that they are homeomorphisms. Let $U \subset \mathbb{C}^{d-1} - \{0\}$ be open. Then $\phi_i^{-1}(U)$ is open because $V = \{0 \neq (z_1, \dots, z_d) \in \mathbb{C}^d : z_i^{-1}(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_d) \in U\}$ is open in $\mathbb{C}^d$, since $\pi V = \phi_i^{-1}(U)$. Let $U \subset \mathbb{C}P^{d-1}$ be open. So $U = \pi V$ for some open $V \subset \mathbb{C}^d$, and $\phi_i(U) = \phi_i(\pi V) = \{z_i^{-1}(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_d) : (z_1, \dots, z_d) \in V\}$, which is open. Hence $\phi_j(U)$ is open in $\mathbb{C}^{d-1}$.

For the inverse, we can write

$$\phi_i^{-1}(z_1, \dots, z_{d-1}) = [z_1, \dots, z_{i-1}, 1, z_{i+1}, \dots, z_{d-1}],$$

and then see that the co-ordinate change functions $\phi_i \circ \phi_j^{-1} : \mathbb{C}^{d-1} - \{0\} \to \mathbb{C}^{d-1} - \{0\}$ are holomorphic.

So $\mathbb{C}P^{d-1}$ is a complex manifold of dimension $d - 1$. We are interested in an open sub-manifold. Let $\mathbb{R}_+$ denote the strictly positive reals. We embed $\mathbb{R}_+^d$ in $\mathbb{C}^d$ in the obvious way. Define an open set

$$Q = \{(z_1, \dots, z_d) : z_j = x_j + iy_j \in \mathbb{C}, x_j, y_j > 0, j = 1, \dots, d\} = \mathbb{R}_+^d + i\mathbb{R}_+^d,$$

so πQ gives an open sub-manifold which we call M. The family of positive matrices $(A_\alpha)_{\alpha \in I}$ act naturally on this manifold. We may define functions

$$A_\alpha : M \to M; \quad A_\alpha([z_1, \dots, z_d]) = [A_\alpha(z_1, \dots, z_d)].$$

Since $A_\alpha \lambda z = \lambda A_\alpha z$ when $z \in \mathbb{C}^d$ and $\lambda \in \mathbb{C}$, and since it is readily seen that $A_\alpha Q \subset Q$, we have that these are well-defined functions on M.

Note that $M \subset U_i$ for each i, so an atlas for M need only contain one map, so we may define

$$\phi : M \to \mathbb{C}^{d-1}; \quad \phi[z_1, \dots, z_d] = z_1^{-1}(z_2, \dots, z_d),$$

which is injective, so we'd like to know what the image $\phi(M)$ is. If $\underline{w} = (w_1, \dots, w_{d-1}) = \phi([r_1 e^{i\theta_1}, \dots, r_d e^{i\theta_d}])$ for some $r_i > 0$ and $\theta_i \in (0, \pi/2)$ for each $i = 1, \dots, d$, then $\arg w_j = \theta_{j+1} - \theta_1 \in (-\pi/2, \pi/2)$ and $\arg w_i - \arg w_j = \theta_{i+1} - \theta_{j+1} \in (-\pi/2, \pi/2)$. These conditions are equivalent to $\mathfrak{Re}\, w_j > 0$ and $\mathfrak{Re}\, w_i \overline{w_j} > 0$. Define

$$D = \{(z_1, \dots, z_{d-1}) \in \mathbb{C}^{d-1} \,:\, \mathfrak{Re}\, z_i > 0 \text{ and } \min_{i \neq j} \mathfrak{Re}\, z_i \overline{z_j} > 0\}$$

and let $(w_1, \dots, w_{d-1}) \in D$. Let $\theta = -\min_i \arg w_i + \epsilon$ for some small $\epsilon > 0$. If $\theta > 0$ then set instead $\theta = \epsilon$. Set $\underline{z} = e^{i\theta}(1, z_1, \dots, z_{d-1})$. We have that $[\underline{z}] \in M$ and that any other pre-image is equal to $[\underline{z}]$. So henceforth we write $\phi : M \to D$, upgrading ϕ to a bijection.

The set D is open, convex, and contains the positive real axis

$$\mathbb{R}_+^{d-1} = \{\underline{z} \in \mathbb{C}^{d-1} \,:\, \mathfrak{Re}\, z_i > 0,\, \mathfrak{Im}\, z_i = 0\}. \tag{3.2}$$

This suggests there are links with the complex cones of Rugh, see (50) for example, in which case far more could be said of the maps we define on this set.

The maps $\phi \circ A_\alpha \circ \phi^{-1} \,:\, D \to D$ have simple expressions and are holomorphic in each variable separately by the algebra of holomorphic functions.

Definition 3.1. For $\alpha \in I$, we define the map $T_\alpha \,:\, D \to D$ by

$$T_\alpha(z) = \phi \circ A_\alpha \circ \phi^{-1}(z)$$

To be explicit, if A_α has entries $a_{ij}^{(\alpha)}$, for $\underline{z} = (z_1, \dots, z_{d-1})$ we can write

$$T_\alpha(\underline{z}) = \left(\frac{a_{21}^{(\alpha)} + a_{22}^{(\alpha)} z_1 + \cdots + a_{2d}^{(\alpha)} z_{d-1}}{a_{11}^{(\alpha)} + a_{12}^{(\alpha)} z_1 + \cdots + a_{1d}^{(\alpha)} z_{d-1}}, \dots, \frac{a_{d1}^{(\alpha)} + a_{d2}^{(\alpha)} z_1 + \cdots + a_{dd}^{(\alpha)} z_{d-1}}{a_{11}^{(\alpha)} + a_{12}^{(\alpha)} z_1 + \cdots + a_{1d}^{(\alpha)} z_{d-1}} \right).$$

For $\underline{\alpha} \in I^n$ we define $T_{\underline{\alpha}} = T_{\alpha_1} \circ \cdots \circ T_{\alpha_n}$.

We remark that in looking at this expression directly, it may not be immediately clear that it maps $D \to D$. These maps have useful properties and correspond well with the original matrices, and similar maps are used in multidimensional continued fraction theory, see the book of Schweiger (54). Similar maps can be used to prove the Perron-Frobenius theorem for finite matrices (see (2) for a proof), but this variation is so that the maps are holomorphic.

It is sometimes useful to define the linear fractional transformation for an arbitrary $d \times d$ matrix B, not just ones with positive entries. Where it makes sense, we define $T_B : \mathbb{C}^{d-1} \to \mathbb{C}^{d-1}$ by

$$T_B(z) = \left(\frac{A_2 \cdot \tilde{z}}{A_1 \cdot \tilde{z}}, \dots, \frac{A_d \cdot \tilde{z}}{A_1 \cdot \tilde{z}} \right),$$

where A_i is the vector given by the ith row of the matrix B, and $\tilde{z}$ is the vector in $\mathbb{C}^d$ given by $(1, z_1, \dots, z_{d-1})$.

Example 3.2. We consider the complex linear fractional transformation associated with the positive 2×2 matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. The transformation is then the map $T : D \to D$ where $D = \{z \in \mathbb{C} : \mathfrak{Re}\, z > 0\}$ given by

$$T(z) = \frac{c + dz}{a + bz}.$$

The boundary of the image $T(D)$ consists of points $T(is)$ for $s \in \mathbb{R}$ (where we have extended the domain of T to include points on the imaginary axis).

Setting $x = \mathfrak{Re}\, T(is)$ and $y = \mathfrak{Im}\, T(is)$, finding the real and imaginary parts of

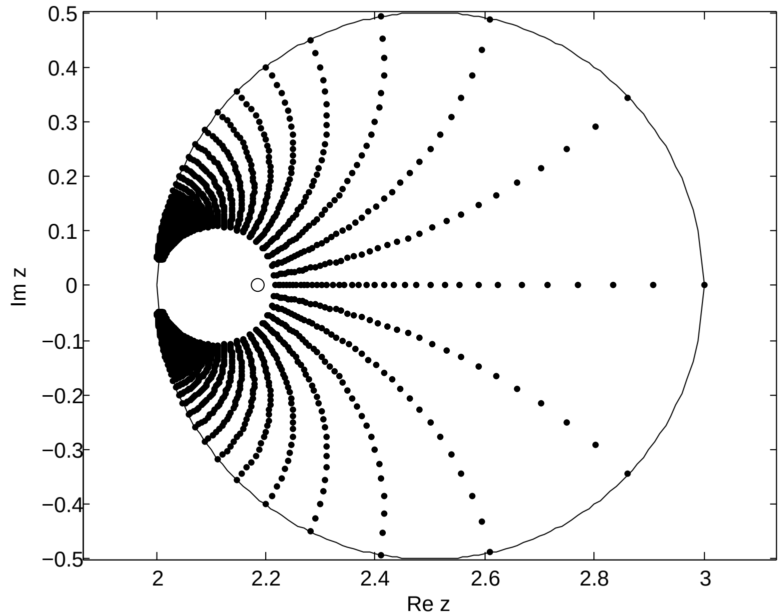


Figure 3.2: A plot of $T(z)$ for various (regularly spaced) $z \in D$ for a 2×2 matrix in example 3.2. The small circle at $z = 2.18$ is the fixed point.

T, finding s^2 in terms of x and substituting into y^2, we arrive at

$$\left(x - \frac{bc + ad}{2ab}\right)^2 + y^2 = \frac{(bc - ad)^2}{4a^2b^2},$$

so $T(D)$ is always the interior of a circle centred on the real axis, and contained entirely in the positive imaginary half of the complex plane. In figure 3.2, we plot the images of regularly spaced points in $[0, 1.8] + i[-10, 10] \subset D$ for the matrix $\left(\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right)$, and include the bounding circle, and the fixed point.

We now need to state and prove some important properties of these maps before we employ them to define transfer operators.

Lemma 3.3. *For all $\alpha \in I$, the map T_α associated with the matrix A_α has $\overline{T_\alpha(D)} \subset D$. In fact,*

$$\overline{\cup_{\alpha \in I} T_\alpha(D)} \subset D \tag{3.3}$$

and this set is compact in D.

Proof. First we wish to show that $\cup_{\alpha \in I} T_\alpha(D)$ is bounded for an arbitrary family of $(d \times d)$-dimensional positive matrices $\{A_\alpha\}_{\alpha \in I}$ where I is a compact set, and the entry for the ith row and jth column of A_α is denoted by $a_{ij}^{(\alpha)}$. We show this by induction on d. For $d = 2$, $\|T_\alpha\|_\infty = \max(|a_{11}^{(\alpha)}/a_{21}^{(\alpha)}|, |a_{12}^{(\alpha)}/a_{22}^{(\alpha)}|)$ by a routine calculation, and this right hand side is uniformly bounded because it is a continuous function of α on a compact set. This establishes the basis for the induction.

Assume now $d > 2$, and the result holds for $d-1$. If the result does not hold for d, there exists a sequence $(y_n)_{n \in \mathbb{N}}$ in $\cup_\alpha T_\alpha(D)$ where $y_n = T_{\alpha_n}(x_n)$ for some sequences $(x_n)_{n\in\mathbb{N}}$ and $(\alpha_n)_{n\in\mathbb{N}}$ in D and I respectively, where some component j of y_n is unbounded. Write

$$y_n^{(j)} = \frac{A_n + a_{jd}^{(\alpha_n)} x_n^{(d-1)}}{B_n + a_{1d}^{(\alpha_n)} x_n^{(d-1)}}$$

where $A_n = a_{j1}^{(\alpha_n)} + a_{j2}^{(\alpha_n)} x_n^{(1)} + \cdots + a_{j(d-1)}^{(\alpha_n)} x_n^{(d-2)}$ and $B_n = a_{11}^{(\alpha_n)} + a_{12}^{(\alpha_n)} x_n^{(1)} + \cdots + a_{1(d-1)}^{(\alpha_n)} x_n^{(d-2)}$. Hence $y_n^{(j)} = f_n(x_n^{(d-1)})$ where $f_n : D \to \mathbb{C}$ is defined by $f_n(x) = (A_n + a_{jd}^{(\alpha_n)} x)/(B_n + a_{1d}^{(\alpha_n)} x)$. But $\|f_n\|_\infty \le \max\{|A_n|/|B_n|, |a_{jd}^{(\alpha_n)}|/|a_{1d}^{(\alpha_n)}|\}$. Since A_n/B_n looks like a fractional transformation of a $(d-1) \times (d-1)$-dimensional family, by induction it is bounded, by M say. Also $\alpha \mapsto |a_{jd}^{(\alpha)}|/|a_{1d}^{(\alpha)}|$ is a continuous function on a compact space I, hence bounded. Hence $(f_n)_{n\in\mathbb{N}}$ is uniformly bounded, and therefore so is $y_n^{(j)}$, a contradiction.

We next show that $\overline{\cup_{\alpha \in I} T_\alpha(D)} \subset D$. Let $z_n \in \cup_{\alpha\in I} T_\alpha(D)$ such that $z_n \to z$ for some $z \in \overline{D}$. Since $T_\alpha(D) = \phi \circ A_\alpha \circ \phi^{-1}(D) = \phi \circ A_\alpha(M)$, we may write $z_n = \phi(A_{\alpha_n}(w_n))$ where $w_n \in M$. Write $w_n = [x_n + iy_n]$ where $x_n, y_n \in \mathbb{R}_+^d$ have positive co-ordinates. Picking a component j of z_n, we see that

$$z_n^{(j)} = \frac{A_{j+1}^{(\alpha_n)} \cdot x_n + i A_{j+1}^{(\alpha_n)} \cdot y_n}{A_1^{(\alpha_n)} \cdot x_n + i A_1^{(\alpha_n)} \cdot y_n},$$

where $A_i^{(\alpha_n)}$ is the vector given by the ith row of A_{α_n}, and hence

$$\begin{aligned}\mathfrak{Re}\, z_n^{(j)} &= \frac{(A_{j+1}^{(\alpha_n)} \cdot x_n)(A_1^{(\alpha_n)} \cdot x_n) + (A_{j+1}^{(\alpha_n)} \cdot y_n)(A_1^{(\alpha_n)} \cdot y_n)}{(A_1^{(\alpha_n)} \cdot x_n)^2 + (A_1^{(\alpha_n)} \cdot y_n)^2} \\ &\geq \frac{(\min_i a_{j+1,i}^{(\alpha_n)})(\min_i a_{1,i}^{(\alpha_n)})(\|x_n\|_1^2 + \|y_n\|_1^2)}{\|A_1^{(\alpha_n)}\|^2(\|x_n\|^2 + \|y_n\|^2)} \\ &\geq \frac{(\min_i a_{j+1,i}^{(\alpha_n)})(\min_i a_{1,i}^{(\alpha_n)})k^2}{\|A_1^{(\alpha_n)}\|^2} \geq B,\end{aligned}$$

where $B > 0$ does not depend on n, and the second to last inequality is because $\|\cdot\|$ and $\|\cdot\|_1$ are equivalent norms giving $\|\cdot\|_1 \geq k\|\cdot\|$. The last inequality is because $\alpha \mapsto (\min_i a_{j+1,i}^{(\alpha)})(\min_i a_{1,i}^{(\alpha)})/\|A_1^{(\alpha)}\|^2$ is a strictly positive continuous function on a compact set, hence its minimum B is strictly positive. Therefore $\mathfrak{Re} \lim_{n\to\infty} z_n^{(j)} \geq B > 0$.

It only remains to show that for any components i, j we have $\mathfrak{Re}\, z^{(i)}\overline{z^{(j)}} > 0$. This works in a very similar fashion to the previous paragraph, and we use the same notation.

$$\begin{aligned}\mathfrak{Re}\, z^{(i)}\overline{z^{(j)}} &= \frac{\mathfrak{Re}(A_{i+1}^{(\alpha_n)} \cdot w_n)\overline{(A_{j+1}^{(\alpha_n)} \cdot w_n)}}{|A_1^{(\alpha_n)} \cdot w_n|^2} \\ &\geq \frac{(\min_k a_{i+1,k}^{(\alpha_n)})(\min_k a_{j+1,k}^{(\alpha_n)})(\|x_n\|_1^2 + \|y_n\|_1^2)}{\|A_1^{(\alpha_n)}\|^2(\|x_n\|^2 + \|y_n\|^2)} \geq B > 0.\end{aligned}$$

This completes the proof that $z \in D$. □

The fixed points of T_α tell us the eigenvectors of A_α and vice versa. The derivative of T_α evaluated at the fixed point tells us about the spectral radius of A_α. The next lemma gives the details of these relationships. We anticipate this result is already in the literature about linear fractional transformations, but have not been able to find a reference.

Lemma 3.4. *Fix any $\alpha \in I$, and let λ_α be the maximal positive eigenvalue for A_α by the Perron-Frobenius Theorem. Let $X_\alpha = (X_\alpha^{(1)}, \ldots, X_\alpha^{(d)})$ be its corresponding* *eigenvector. Then* $x_\alpha = (X_\alpha^{(2)}/X_\alpha^{(1)}, \ldots, X_\alpha^{(d)}/X_\alpha^{(1)})$ *is the unique fixed point of* T_α.

We can recover information about the maximal eigenvalue from the fractional transformation using the formula

$$\lambda_\alpha = \left(\frac{\det A_\alpha}{\det DT_\alpha(x_\alpha)}\right)^{\frac{1}{d}}$$

Proof. We have

$$\begin{aligned} T_\alpha(x_\alpha) &= \left(\frac{a_{21}^{(\alpha)}X_\alpha^{(1)} + \cdots + a_{2d}^{(\alpha)}X_\alpha^{(d)}}{a_{11}^{(\alpha)}X_\alpha^{(1)} + \cdots + a_{1d}^{(\alpha)}X_\alpha^{(d)}}, \ldots, \frac{a_{d1}^{(\alpha)}X_\alpha^{(1)} + \cdots + a_{dd}^{(\alpha)}X_\alpha^{(d)}}{a_{11}^{(\alpha)}X_\alpha^{(1)} + \cdots + a_{1d}^{(\alpha)}X_\alpha^{(d)}}\right) \\ &= \left(\frac{(A_\alpha X_\alpha)_2}{(A_\alpha X_\alpha)_1}, \ldots, \frac{(A_\alpha X_\alpha)_d}{(A_\alpha X_\alpha)_1}\right) = \left(\frac{\lambda_\alpha X_2}{\lambda_\alpha X_1}, \ldots, \frac{\lambda_\alpha X_d}{\lambda_\alpha X_1}\right) = x_\alpha. \end{aligned}$$

That it is unique follows from a version (55) of the Perron-Frobenius theorem which tells us that the only positive eigenvector is the one corresponding to the maximal positive eigenvalue. This gives uniqueness because if y is another fixed point, then $(1, y_1, \ldots, y_{d-1})$ would be another positive eigenvector.

To prove the formula for the eigenvalue, we fix α and write $A_\alpha = QJQ^{-1}$ where J is the block Jordan form matrix for A_α and Q is the associated change of basis matrix. This decomposition is not unique, because we can re-order the blocks in J and this permutes columns of Q. Consider the term $T_{Q^{-1}}(x_\alpha)$. If the first column of Q is X_α, we have

$$T_{Q^{-1}}(x_\alpha) = \phi \circ Q^{-1} \circ \phi^{-1}(x_\alpha) = \phi(X_\alpha^{(1)} Q^{-1}(X_\alpha)) = (0, \ldots, 0),$$

and if X_α is not the first column of Q then the term is undefined. Since this term occurs in the subsequent equations, we require that the ordering of the decomposition has X_α as the first column, and therefore has λ_α as the first element of the Jordon block. The rest of the ordering does not matter.

We note that $T_\alpha = T_{A_\alpha}|_D = (T_Q \circ T_J \circ T_{Q^{-1}})|_D$ and that $T_J(T_{Q^{-1}}(x_\alpha)) = T_{Q^{-1}}(x_\alpha)$, hence

$$\begin{aligned} DT_\alpha(x_\alpha) &= D(T_Q \circ T_J \circ T_{Q^{-1}})(x_\alpha) \\ &= DT_Q(T_J \circ T_{Q^{-1}} x_\alpha) DT_J(T_{Q^{-1}} x_\alpha) DT_{Q^{-1}}(x_\alpha) \\ &= DT_Q(T_{Q^{-1}} x_\alpha) DT_J(T_{Q^{-1}} x_\alpha) DT_{Q^{-1}}(x_\alpha), \end{aligned}$$

and taking the determinant, rearranging, then unapplying the chain rule for Jacobians, we have

$$\begin{aligned} \det DT_\alpha(x_\alpha) &= \det(DT_Q(T_{Q^{-1}} x_\alpha) DT_{Q^{-1}}(x_\alpha)) \det DT_J(T_{Q^{-1}} x_\alpha) \\ &= \det DT_J(T_{Q^{-1}} x_\alpha). \end{aligned}$$

We denote the diagonal of J as $(\lambda_1, \ldots, \lambda_d)$ with $\lambda_1 = \lambda_\alpha$ and the upper diagonal as $(\rho_1, \ldots, \rho_{d-1})$ with each ρ_i equalling 0 or 1. Since λ_1 is in a 1×1 block, $\rho_1 = 0$. A calculation gives that $\det DT_J(T_{Q^{-1}} x_\alpha) = \lambda_1^{-(d-1)} \prod_{i=2}^d \lambda_i$. Hence we have

$$\left(\frac{\det A_\alpha}{\det DT_\alpha(x_\alpha)} \right)^{\frac{1}{d}} = \left(\frac{\det J}{\det DT_J(T_{Q^{-1}} x_\alpha)} \right)^{\frac{1}{d}} = \left(\frac{\prod_{i=1}^d \lambda_i}{\lambda_1^{-(d-1)} \prod_{i=2}^d \lambda_i} \right)^{\frac{1}{d}} = \lambda_1,$$

completing the proof. □

Note that the previous lemma also works for $\underline{\alpha} \in I^n$ in place of $\alpha \in I$.

If we restrict our attention to the family $(T_\alpha)_{\alpha \in I}$ on a suitable compact set, which contains the fixed points, then we retain information about the spectral radii of the matrices. We first need a standard result from point set topology.

Lemma 3.5. *Let F be a compact set, and U be open, such that $F \subset U$. Then there exists an open bounded set V such that $F \subset V$ and $\overline{V} \subset U$. In addition if U is convex then V can be made to be convex.*

Proof. For the first part of the proof, see for example Theorem 2.7 in the book of

Rudin (46), which applies to any locally compact Hausdorff topological space. This gives us an open bounded set V satisfying $F \subset V \subset \overline{V} \subset U$. To obtain convexity, we replace V by its convex closure V'. This is still open, and certainly $F \subset V' \subset U$. If $x \in \overline{V}'$ then $x = \lim_{n\to\infty} t_n a_n + (1-t_n) b_n$ with $t_n \in [0,1]$, $a_n \in V$, and $b_n \in V$. We may pass to subsequences to assume $t_n \to t \in [0,1]$, $a_n \to a \in \overline{V} \subset U$ and $b_n \to b \in \overline{V} \subset U$. Hence $x = ta + (1-t)b$, and since U is convex, $x \in U$, so $\overline{V}' \subset U$. $\square$

The convexity ensures that the sets this lemma produces are connected. Now $\overline{\cup_{\alpha\in I} T_\alpha(D)}$ is compact by lemma 3.3. So apply lemma 3.5, to get an open, bounded, connected set U such that $\overline{\cup_{\alpha\in I} T_\alpha(D)} \subset U \subset \overline{U} \subset D$. Now we have $\overline{\cup_{\alpha\in I} T_\alpha(U)} \subset \overline{\cup_{\alpha\in I} T_\alpha(D)} \subset U$, so we can now take the maps T_α to be defined $T_\alpha : U \to U$. Note that U still contains all of the fixed points, because if $T_\alpha x_\alpha = x_\alpha$ then $x_\alpha \in \overline{\cup_{\alpha\in I} T_\alpha(D)} \subset U$. Because D is a large set, we have a large choice for the set U. It might be possible to simplify subsequent work by choosing it to be a suitable polydisc. This is certainly possible for 2×2 matrices.

Lemma 3.6. *There exists an $\epsilon > 0$ such that $\|T_\alpha(x) - y\| \geq \epsilon$ for all $\alpha \in I$, $x \in U$ and $y \in \mathbb{C}^{d-1} \setminus U$.*

Proof. Set $K = \overline{\cup_{\alpha\in I} T_\alpha U}$ which is compact by lemma 3.3, and $K \subset U$. For each $x \in K$, there exists $\delta_x > 0$ such that $B(x, 2\delta_x) \subset U$. The compactness of K gives the existence of a finite subset $\{x_1, \ldots, x_N\} \subset K$ such that $K \subset \cup_{i=1}^N B(x_i, \delta_{x_i})$. Fix $x \in U$, $\alpha \in I$, $y \in \mathbb{C}^{d-1} \setminus U$. $T_\alpha(x) \in K$ so $T_\alpha(x) \in B(x_i, \delta_{x_i})$ for some i. $y \notin B(x_i, 2\delta_{x_i})$ so $2\delta_{x_i} \leq \|x_i - y\| \leq \|x_i - T_\alpha(x)\| + \|T_\alpha(x) - y\| \leq \delta_{x_i} + \|T_\alpha(x) - y\|$ implying $\delta_{x_i} \leq \|T_\alpha(x) - y\|$. The proof of this lemma is completed by setting $\epsilon = \min\{\delta_{x_i} \mid i = 1, \ldots, N\}$. $\square$

We can use these results to enable us to find an open set on which we can define the transfer operator, and be able to write down an explicit formula for its trace.

We will also require a suitable metric which makes $\{T_\alpha\}_{\alpha \in I}$ a family of uniform contractions.

Lemma 3.7. *There exists a metric d on U, and a constant $0 < k < 1$ such that*

$$d(T_\alpha x, T_\alpha y) \leq k d(x, y) \tag{3.4}$$

for all $\alpha \in I$ and all $x, y \in D$. There exists a constant $m > 0$ such that

$$\|x - y\| \leq m d(x, y) \tag{3.5}$$

for all $x, y \in U$. (As usual $\|\cdot\|$ is the Euclidean norm in $\mathbb{C}^{d-1}$). For each $x \in U$, there exists a constant $r > 0$ and a constant $M_x > 0$ such that

$$d(x, y) \leq M_x \|x - y\| \tag{3.6}$$

for all $y \in B_d(x, r)$.

Proof. The metric is the Carathéodory-Reiffen Finsler metric used in the paper (10) of Earle and Hamilton to prove their fixed point theorem. We give an outline of the construction used in the paper, and add in detail to show the uniformity of the contraction constant for our maps, and the existence of the local bound to the Euclidean metric.

Let $H^\infty(U)$ denote the bounded holomorphic functions on U. Define

$$\alpha(x, v) = \sup\{\|Df(x)v\| \, : \, \|f\|_\infty \leq 1, f \in H^\infty(U)\}$$

for all $x \in U$ and $v \in \mathbb{C}^{d-1}$. Let Γ denote the set of curves $[0, 1] \to U$ with piecewise continuous derivatives. For $\gamma \in \Gamma$, set

$$L_\alpha(\gamma) = \int_0^1 \alpha(\gamma(t), \gamma'(t)) dt,$$

and define

$$d(x,y) = \inf\{L_\alpha(\gamma) \,:\, \gamma \in \Gamma, \gamma(0) = x, \gamma(1) = y\}.$$

It is straightforward to establish that this is a pseudometric, and inequality 3.5 will show that it is a metric. For any holomorphic $g : U \to \mathbb{C}^{d-1}$ such that $g(U) \subset U$ the paper proves that $\alpha(g(x), Dg(x)v) \leq \alpha(x,v)$ for all $x \in U$ and $v \in \mathbb{C}^{d-1}$. Choose $\epsilon > 0$ by 3.6 such that $\|T_\beta(x) - y\| > \epsilon$ for all $x \in U$, $y \notin U$ and $\beta \in I$, and choose $D > 0$ such that $\|x\| \leq D$ for all $x \in U$. Set $\mu = \epsilon/2D$. Fix $x \in U$ and define $g_\beta : U \to \mathbb{C}^{d-1}$ by $g_\beta(y) = (1+\mu)T_\beta(y) - \mu T_\beta(x)$ for all $y \in U$. Since $T_\beta(x) = g_\beta(x)$ and $Dg_\beta(x) = (1+\mu)DT_\beta(x)$, we have $\alpha(T_\beta(x), DT_\beta(x)v) = (1+\mu)^{-1}\alpha(g_\beta(x), Dg_\beta(x)v)$ for all $x \in U$, $v \in \mathbb{C}^{d-1}$ and $\beta \in I$. Hence $\alpha(T_\beta(x), DT_\beta(x)v) \leq (1+\mu)^{-1}\alpha(x,v)$, so integrating gives inequality 3.4 where $k = (1+\mu)^{-1} < 1$.

To show inequality 3.5, fix $x \in U$ and set $g(y) = m\|x - y\|$ where $\|x\| < m^{-1}$ for all $x \in U$. Then $Dg(x)v = m\|v\|$, so $\alpha(x,v) \geq m\|v\|$ for all $x \in U$, $v \in \mathbb{C}^{d-1}$. So $L_\alpha(\gamma) \geq m\int_0^1 \|\gamma'(t)\|dt \geq m\|x-y\|$ for all curves $\gamma \in \Gamma$ from $x \in U$ to $y \in U$.

Finally, to show inequality 3.6, fix $x \in U$ and since U is d-open (we have that $\|\cdot\|$-open implies d-open) choose $r_0 > 0$ such that $B_d(x, r_0) \subset U$. For an arbitrary $f \in H^\infty(U)$ we may use the Cauchy estimates for several complex variables (see (27), p. 104) to bound partial derivatives on $B_d(x, r_0)$, so that $\|Df(y)\| \leq C_y\|f\|_\infty$ for some $C_y > 0$ (which does not depend on f) and varies continuously with $y \in B_d(x, r_0)$. Take some $\epsilon > 0$ and choose r_1 such that $|C_x - C_y| < \epsilon$ when $d(x,y) < r_1$, and choose $r = \min(r_0, r_1)$ so that $\|Df(y)\| \leq (C_x+\epsilon)\|f\|_\infty$ for all $y \in B_d(x,r)$. Set $M_x = C_x + \epsilon$. Then for $v \in \mathbb{C}^{d-1}$ we have $\|Df(y)v\| \leq \|Df(y)\|\|v\| \leq M_x\|f\|_\infty\|v\|$, therefore $\alpha(y,v) \leq M_x\|v\|$. Writing Γ' for the set of curves $\gamma \in \Gamma$ inside $B_d(x,r)$ with $\gamma(0) = x$ and $\gamma(1) = y$, we have $L_\alpha(\gamma) \leq M_x\int_0^1 \|\gamma'(t)\|dt$ on such curves, and taking infimums of both sides gives inequality 3.6, since $\inf\{\int_0^1 \|\gamma'(t)\|dt \,:\, \gamma \in \Gamma'\} = \|x-y\|$. □

This lemma states that d gives the same topology as the Euclidean metric, and is complete.

Remark 3.8. The existence of the unique fixed points for the maps T_α could be obtained using the previous lemma, instead of using the Perron-Frobenius theorem. The conjugate $\overline{x_\alpha}$ is also a fixed point, hence $\overline{x_\alpha} = x_\alpha$ by uniqueness and so each component is real. More generally, from a numerical point of view, it is perhaps interesting that iterating a linear fractional transformation associated with a strictly positive matrix converges on a value, from which we can easily recover the maximal positive eigenvector (and hence spectral radius). Perhaps the CRF metric could be used for a proof of the Perron-Frobenius theorem in a similar way to Birkhoff's proof using the Hilbert Projective metric, for which the action of positive matrices on the positive quadrant of $\mathbb{R}P^d$ is a contraction.

We can use lemma 3.7 to define the family $\{T_\alpha\}_{\alpha\in I}$ on smaller connected sets that still contain the fixed points and satisfy equation 3.3, so that we can keep the information we need but make the derivatives behave nicely. Recovering the spectral radius involves taking a complex log of the derivative, so we need a domain where the derivative behaves nicely. Recall that D contains the positive real axis $\mathbb{R}_+^{d-1} = \{z \in \mathbb{C}^{d-1} : \mathfrak{Re}\, z_i > 0, \mathfrak{Im}\, z_i = 0\}$.

Lemma 3.9. *Let $\delta > 0$. Define*

$$V_\delta = \{x \in U \,:\, \exists y \in U \cap \mathbb{R}_+^{d-1},\, d(x,y) \leq \delta\}.$$

Then V_δ is open, connected, contains the fixed points, and satisfies equation 3.3, i.e. $\overline{\cup_{\alpha\in I} T_\alpha(V_\delta)} \subset V_\delta$. We can also replace V_δ with a smaller set which has all the same properties but is bounded with respect to d.

Proof. First we fix $\alpha \in I$ and show $T_\alpha(V_\delta) \subset V_{\delta(k+1)/2}$, where $k < 1$ is the uniform contraction ratio from 3.7. Let $x \in V_\delta$. There exists $y \in U \cap \mathbb{R}_+^{d-1}$ with $d(x,y) < \delta$. Now $d(T_\alpha(x), T_\alpha(y)) \leq kd(x,y) \leq k\delta < \delta(k+1)/2$, and since T_α maps points of $\mathbb{R}_+^{d-1}$ to points of $\mathbb{R}_+^{d-1}$, we have $T_\alpha(y) \in U \cap \mathbb{R}_+^{d-1}$. Hence $T_\alpha(x) \in V_{\delta(k+1)/2}$. This gives

$\overline{\cup_{\alpha\in I}T_\alpha(V_\delta)} \subset \overline{V_{\delta(k+1)/2}} \subset V_\delta$. Clearly V_δ is open and connected, in fact it is convex because it is the intersection of a convex set with the δ-neighbourhood of another convex set. V_δ contains the fixed points because it contains $\mathbb{R}_+^{d-1} \cap U$. Finally, to get a smaller set which is bounded with respect to d, we use lemma 3.5 to find a convex open bounded set V'_δ such that $\overline{\cup_{\alpha\in I}T_\alpha(V_\delta)} \subset V'_\delta \subset \overline{V'_\delta} \subset V_\delta$. Again this contains fixed points: $x_\alpha \in \overline{\cup_{\alpha\in I}T_\alpha(V_\delta)} \subset V'_\delta$. We have $\overline{\cup_{\alpha\in I}T_\alpha(V'_\delta)} \subset \overline{\cup_{\alpha\in I}T_\alpha(V_\delta)} \subset V'_\delta$. Since $\overline{V'_\delta}$ is compact, a compactness argument shows that V'_δ is bounded. □

Because of the relationship between the spectral radius of a matrix and the derivative of the associated fractional transformation at the fixed point, we will study the function defined by:

$$f_\alpha(z) = -\log(\det DT_\alpha(z))/d, \tag{3.7}$$

so that $f_\alpha(x_\alpha) = \log \lambda_\alpha$. The danger with this definition is that we take the log of a complex valued function, but fortunately we can define this function on a suitably small set, using the next lemma, which uses lemma 3.9 to construct sets arbitrarily close to the interval on the real line containing the fixed points. The derivative then behaves nicely enough to take the complex logarithm.

Alternatively, we could take the absolute value of the function inside the log, which means the function isn't holomorphic unless we consider it as a function of $z = x + iy$ and complexify in each of x and y, following Remark 2.14. However, we would then have other work to do to show that the new versions of T_α map inside the new domain, and we would have slower convergence.

Lemma 3.10. *Fix any $\alpha \in I$. There exists a set V such that the function $f_\alpha : V \to \mathbb{C}$ given by equation 3.7 is well-defined and in $F(V)$. Furthermore, the set V is open, connected, bounded, and has $\overline{\cup_{\alpha\in I}T_\alpha(V)} \subset V$.*

Proof. First note that for $x \in U \cap \mathbb{R}_+^{d-1}$, $\det DT_\alpha(x)$ is real and positive for all

$\alpha \in I$. If not, there would be a point where $\det DT_\alpha(x) = 0$ contradicting the inverse function theorem, because T_α is locally invertible. Therefore there exists an open neighbourhood W of $U \cap \mathbb{R}_+^{d-1}$ such that $\mathfrak{Re} \det DT_\alpha(z) > 0$ for all $z \in W$ and $\alpha \in I$. Using the set V_δ from lemma 3.9, choose $\delta > 0$ such that $V_\delta \subset W$. Then V_δ has all the required properties, and since $\det DT_\alpha(x)$ avoids $(-\infty, 0]$, we may take the complex log. Letting $V = V_\delta$ proves the lemma. $\square$

We may now assume the family of functions $\{T_\alpha\}_{\alpha \in I}$ is defined on $V \to V$. We also extend the definition of the function f_α and write

$$f_{\underline{\alpha}}(z) = -\log(\det DT_{\underline{\alpha}}(z))/d,$$

for $z \in V$ and any $\underline{\alpha} \in I^n$ for any n.

We would like the function $f_{\underline{\alpha}}$ to have bounded variation over all of V and all sequences $\underline{\alpha} \in I^n$. Since we have the log of a (determinant of a) derivative, we can employ the usual trick of applying the chain rule, then using the following simple lemma.

Lemma 3.11. *There exists constants $A > 0$, $0 < k < 1$ such that for any sequence $(\alpha_n)_{n\in\mathbb{N}} \in I$ and any $x, y \in V$ we have that for all n*

$$\|T_{\alpha_1\cdots\alpha_n}(x) - T_{\alpha_1\cdots\alpha_n}(y)\| < Ak^n.$$

Proof. Fix a sequence $(\alpha_n)_{n\in\mathbb{N}} \in I$, $x, y \in V$, and $n \in \mathbb{N}$. Then using k and m from lemma 3.7,

$$\begin{aligned}\|T_{\alpha_1\cdots\alpha_n}(x) - T_{\alpha_1\cdots\alpha_n}(y)\| &\leq m^{-1} d(T_{\alpha_1} \circ \cdots \circ T_{\alpha_n}(x), T_{\alpha_1} \circ \cdots \circ T_{\alpha_n}(y)) \\ &\leq m^{-1} k^n d(x, y).\end{aligned}$$

Since $d(x, y) \leq M$ for some $M > 0$ on V, setting $A = M/m$ proves the result. $\square$

This gives rise to the following result about $f_{\underline{\alpha}}$.

Lemma 3.12. *There exists a constant B such that for any $x, y \in V$ and any $\underline{\alpha} \in I^n$ for any $n \in \mathbb{N}$, we have*

$$|f_{\underline{\alpha}}(x) - f_{\underline{\alpha}}(y)| \leq B$$

Proof. By applying the chain rule we have for any $\underline{\alpha} \in I^n$

$$f_{\underline{\alpha}}(x) = \sum_{m=1}^{n} f_{\alpha_m}(T_{\alpha_{m+1}} \circ \cdots \circ T_{\alpha_n}(x)).$$

Furthermore each f_α ($\alpha \in I$) is Lipschitz with constant b, i.e. for any $x, y \in V$, $|f_\alpha(x) - f_\alpha(y)| \leq b\|x - y\|$. Hence, by lemma 3.11,

$$\begin{aligned} |f_{\underline{\alpha}}(x) - f_{\underline{\alpha}}(y)| &\leq b \sum_{m=1}^{n} \|T_{\alpha_{m+1}} \circ \cdots \circ T_{\alpha_n}(x) - T_{\alpha_{m+1}} \circ \cdots \circ T_{\alpha_n}(y)\| \\ &\leq bA \sum_{m=1}^{n} k^m < B, \end{aligned}$$

where $B = bA\frac{k}{1-k}$. $\square$

3.3 Transfer Operators

As before, let $F(V)$ denote the Banach space of bounded holomorphic functions on V, equipped with the supremum norm. See definition 2.3 in the background chapter.

Definition 3.13. We define a transfer operator $\mathcal{L} : F(V) \to F(V)$ by

$$\mathcal{L}f(z) = \int_I f(T_\alpha z) d\rho(\alpha),$$

which is a special case of the more general $\mathcal{L}_{\beta,t} : F(V) \to F(V)$ for $\beta \in I$ and $t \in \mathbb{C}$ defined by

$$\mathcal{L}_{\beta,t} f(z) = \int_I f(T_\alpha z) e^{t f_\beta(T_\alpha z)} d\rho(\alpha),$$

with $\mathcal{L}_{\beta,0} = \mathcal{L}$ for any $\beta \in I$.

Lemma 3.14. *The operator $\mathcal{L}_{\beta,t}$ is well-defined for all $t \in \mathbb{C}$ and all $\beta \in I$. Furthermore, the family $\{\mathcal{L}_{\beta,t}\}_{t\in\mathbb{C}}$ is analytic.*

Proof. First fix $\alpha \in I$ and $f \in F(V)$. Then $T_\alpha(\overline{V}) \subset \overline{T_\alpha(V)} \subset V$, so for any $z \in \overline{V}$, $f \circ T_\alpha$ is continuous at z. Similarly for any $z \in V$, $f \circ T_\alpha$ is holomorphic at z. Hence $f \circ T_\alpha \in F(V)$ for all $\alpha \in I$.

Now fix $\beta \in I$, $t \in \mathbb{C}$, and $f \in F(V)$. We want to show $\mathcal{L}_{\beta,t} f \in F(V)$. Let $\gamma : [0,1] \to V$ be a smooth closed curve which keeps all but one of the co-ordinates in V fixed, with piecewise continuous derivative. Then

$$\int_0^1 \int_I |f(T_\alpha\gamma(s))e^{tf_\beta(T_\alpha\gamma(s))}\gamma'(s)|d\rho(\alpha)ds \leq \|f\|_{F(V)}\|e^{tf_\beta}\|_{F(V)}L(\gamma) < \infty$$

where $L(\gamma)$ is the length of γ. Hence Fubini's theorem may be applied to swap the following integrals, and since $f \circ T_\alpha$ is holomorphic, we have

$$\oint_\gamma \mathcal{L}_{\beta,t} f(z)dz = \int_I \int_0^1 f(T_\alpha\gamma(s))e^{tf_\beta(T_\alpha\gamma(s))}\gamma'(s)dsd\rho(\alpha) = 0.$$

So by Morera's theorem (see proposition 2.1), $\mathcal{L}_{\beta,t} f \in \mathrm{Hol}(V)$. It is bounded because $\|\mathcal{L}_{\beta,t} f(z)\|_{F(V)} \leq \|f\|_{F(V)}\|e^{tf_\beta(\cdot)}\|_{F(V)} \leq \infty$.

Again fix $\beta \in I$. To show that the family $\{\mathcal{L}_{\beta,t}\}_{t\in\mathbb{C}}$ is analytic, we again use Morera's theorem. Let $\gamma : [0,1] \to \mathbb{C}$ be a closed smooth curve. We want to show the operator $\oint_\gamma \mathcal{L}_{\beta,z}dz$ is zero, that is for any fixed $w \in F(V)$ and $x \in V$, we have $\left(\oint_\gamma \mathcal{L}_{\beta,z}dz\right) w(x) = 0$. Now swapping the integrals as before, we have

$$\begin{aligned}\left(\oint_\gamma \mathcal{L}_{\beta,z}dz\right) w(x) &= \oint_\gamma \mathcal{L}_{\beta,z}w(x)dz = \oint_\gamma \int_I w(T_\alpha x)e^{zf_\beta(T_\alpha x)}d\rho(\alpha)dz \\ &= \int_I \oint_\gamma w(T_\alpha x)e^{zf_\beta(T_\alpha x)}dz\, d\rho(\alpha) = \int_I 0 d\rho = 0\end{aligned}$$

since $z \mapsto e^{zf_\beta(T_\alpha x)}$ is holomorphic for fixed $x \in V$, $\alpha, \beta \in I$. □

By the work of Bandtlow and Jenkinson (5), which applies for general measures rather than just counting measures, we have that $\mathcal{L}_{\beta,t}$ is strongly nuclear, although we reproduce this general proof here, more or less verbatim.

Theorem 3.15. *For any $t \in \mathbb{C}$ and any $\beta \in I$, the operator $\mathcal{L}_{\beta,t}$ is strongly nuclear.*

Proof. Fix $\beta \in I$ and $t \in \mathbb{R}$. We follow the proof of Proposition 2.10 in (5). Using lemma 3.5, we choose a bounded, connected, open subset V' set that $\overline{\cup_{\alpha\in I} T_\alpha(V)} \subset V'$ and $\overline{V'} \subset V$. We note that the operator $\hat{\mathcal{L}}f = \int_I f \circ T_\alpha \cdot e^{tf_\beta \circ T_\alpha} d\rho(\alpha)$ defines a continuous operator $F(V') \to F(V)$. To see this fix $f \in F(V')$, then

$$\|\hat{\mathcal{L}}f\|_{F(V)} \leq \int_I \|f\|_{F(V')} \|e^{tf_\beta}\|_{F(V)} d\rho = \|f\|_{F(V')} \|e^{tf_\beta}\|_{F(V)} < \infty.$$

We wish to show the embedding $J : F(V) \hookrightarrow F(V')$ is strongly nuclear. For then if $f \in F(V)$, $\mathcal{L}_{\beta,t}f = \hat{\mathcal{L}}Jf \in F(V)$ so $F(V)$ is $\mathcal{L}_{\beta,t}$-invariant, hence $\mathcal{L}_{\beta,t}$ strongly nuclear.

Choose an open, connected, bounded set V'' such that $\overline{V'} \subset V''$ and $\overline{V''} \subset V$. Now $J = J_2 J_1$ where $J_1 : F(V) \to \mathrm{Hol}(V'')$ and $J_2 : \mathrm{Hol}(V'') \to F(V')$. J_2 is bounded. By Grothendieck (13) the Frechet space $\mathrm{Hol}(V'')$ is nuclear so J_2 is p-nuclear for every $p > 0$, so since strongly nuclear operators are an ideal, we only need to show J_1 is continuous.

By the closed graph theorem, it is enough to show that if $f_n \to f$ in $F(V)$ and $J_1 f_n \to g$ in $\mathrm{Hol}(V'')$ then $g = J_1 f$. We have that $f_n(z) \to f(z)$ for all $z \in V''$. Since point evaluation is also continuous on $\mathrm{Hol}(V'')$, so $f_n(z) = J_1 f_n(z) \to g(z)$ as $n \to \infty$ for all $z \in V''$, hence $g = f|_{V''}$. □

We can now prove standard results about the spectrum of $\mathcal{L}$. The argument follows methods used in the books of, for example, Parry and Pollicott (38), and Baladi (2). Here we take advantage of the lack of weight terms, and the compactness established by the nuclearity of $\mathcal{L}$, to make simplifications. More general results in this setting (with weights) can be found in the papers of Mayer (32), and Bandtlow and Jenkinson (4).

Theorem 3.16. *The transfer operator $\mathcal{L}$ satisfies the following:*

1. 1 *is the simple positive maximal eigenvalue for $\mathcal{L}$, with the constant function* 1 *its eigenfunction.*

2. *The rest of the spectrum is contained in a disc of radius strictly less than* 1*.*

3. *$\mathcal{L}$ has an eigenprojection μ, that is $\mathcal{L}^\star \mu = \mu$.*

4. *For any $f \in F(V)$, we have*

$$\lim_{n\to\infty} \mathcal{L}^n f = \mu(f) \tag{3.8}$$

uniformly.

Proof. It is clear that 1 is an eigenvalue with eigenfunction 1. We prove that 1 is a simple eigenvalue by showing 1 is a simple eigenvalue of the dual of $\mathcal{L}$, and this is done by constructing a eigenprojection μ for 1, which shows the third part of the theorem.

Fix $f \in F(V)$. Consider the sequence of functions $(\mathcal{L}^n f)_{n\in\mathbb{N}}$. Since f is Lipschitz with constant $\mathrm{Lip}(f)$, then for $x, y \in V$,

$$|\mathcal{L}^n f(x) - \mathcal{L}^n f(y)| \leq \int_{I^n} \mathrm{Lip}(f)\, \|T_{\underline{\alpha}}(x) - T_{\underline{\alpha}}(y)\| d\rho^n(\underline{\alpha}) \leq \mathrm{Lip}(f)\, \theta^n m d(x,y),$$

hence for $\epsilon > 0$, we can choose $\delta > 0$ such that $|\mathcal{L}^n f(x) - \mathcal{L}^n f(y)| \leq \epsilon$ whenever $\|x - y\| < \delta$ for any $n \in \mathbb{N}$, hence the sequence is equicontinuous. Since $\|\mathcal{L}^n f\|_\infty \leq \|f\|_\infty$ we have that the sequence is bounded. Apply the Arzelà-Ascoli theorem to get a function $f_\star$ on $\overline{V}$ and a subsequence n_k such that $\mathcal{L}^{n_k} f \to f_\star$ uniformly as $k \to \infty$.

We show $f_\star$ is constant. Pick $x, y \in V$ and $\epsilon > 0$. Now

$$\begin{aligned} |f_\star(x) - f_\star(y)| &\leq |f_\star(x) - \mathcal{L}^{n_k} f(x)| + |\mathcal{L}^{n_k} f(x) - \mathcal{L}^{n_k} f(y)| + |\mathcal{L}^{n_k} f(y) - f_\star(y)| \\ &\leq 2\|\mathcal{L}^{n_k} f - f_\star\|_\infty + \int_{I^{n_k}} \mathrm{Lip}(f)\, \|T_{\underline{\alpha}}(x) - T_{\underline{\alpha}}(y)\| d\rho^{n_k}(\underline{\alpha}) \\ &\leq 2\|\mathcal{L}^{n_k} f - f_\star\|_\infty + \mathrm{Lip}(f)\, m\theta^{n_k} d(x, y), \end{aligned}$$

hence if k is large enough such that $\mathrm{Lip}(f)\, m\theta^{n_k} d(x, y) < \epsilon/3$ and $\|\mathcal{L}^{n_k} f - f_\star\|_\infty \leq \epsilon/3$, then $|f_\star(x) - f_\star(y)| < \epsilon$. Since ϵ is arbitrary, $f_\star$ is a constant in $\mathbb{C}$.

In fact $\mathcal{L}^n f \to f_\star$ uniformly. Since $\mathcal{L} f_\star = f_\star$ (since 1 is an eigenvalue of $\mathcal{L}$ and $f_\star$ is constant), and $\|\mathcal{L} f\|_\infty \leq \|f\|_\infty$, we have for any $m < n$,

$$\|\mathcal{L}^n f - f_\star\|_\infty \leq \|\mathcal{L}^m f - f_\star\|_\infty.$$

Then if we pick $\epsilon > 0$ and choose k such that $\|\mathcal{L}^{n_k} f - f_\star\|_\infty < \epsilon$, then whenever $n > n_k$ we have

$$\|\mathcal{L}^n f - f_\star\|_\infty \leq \|\mathcal{L}^{n_k} f - f_\star\|_\infty < \epsilon.$$

Define a bounded linear functional $\mu \in F(V)^\star$ by $\mu(f) = f_\star \in \mathbb{C}$. Since $\mathcal{L}^n(\mathcal{L} f) \to \mathcal{L} f_\star$ and $\mathcal{L} f_\star = f_\star$, we have that $\mu(\mathcal{L} f) = \mu(f)$, that is $\mathcal{L}^\star \mu = \mu$. This shows parts 3 and 4 of the theorem. If there is another eigenprojection $\nu \in F(V)^\star$ such that $\mathcal{L}^\star \nu = \nu$, then

$$|\nu(f) - \nu(f_\star)| = |\nu(\mathcal{L}^n f) - \nu(f_\star)| \leq \|\nu\| \|\mathcal{L}^n f - f_\star\|_\infty \to 0,$$

so $\nu(f) = \nu(f_\star) = \nu(\mu(f) \cdot 1) = \mu(f)\nu(1)$, and hence ν is just a multiple of μ. Hence 1 is a simple eigenvalue of $\mathcal{L}^\star$. Since $\mathcal{L}$ is compact, 1 is also a simple eigenvalue for $\mathcal{L}$.

Now if $\mathcal{L} f = \lambda f$ then $|\lambda| \|f\|_\infty = \|\mathcal{L} f\|_\infty \leq \|f\|_\infty$, so $|\lambda| \leq 1$. If $e^{i\theta}$ is another eigenvalue of unit modulus ($\theta \neq 0$), and $\mathcal{L} f = e^{i\theta} f$ for some $f \neq 0$, then $\mu(f) =$

$\mu(\mathcal{L}f) = \mu(e^{i\theta}f) = e^{i\theta}\mu(f)$ hence $\mu(f) = 0$. But $\|\mathcal{L}^n f\|_\infty = \|e^{in\theta}f\|_\infty = \|f\|_\infty$ for all n. Hence $\mathcal{L}^n f$ cannot tend to $\mu(f) = 0$ uniformly, contradicting part 4 of the theorem. By compactness of the operator, all other eigenvalues must have modulus strictly less than 1. This shows part 2 and completes the proof. □

Part 4 of the previous theorem in particular is key to proving the next result. But first, it is useful to have an explicit expression for $\mathcal{L}^n$,

$$\mathcal{L}^n w(z) = \int_I \cdots \int_I w(T_{\alpha_1} \cdots T_{\alpha_n} z) d\rho(\alpha_1) \cdots d\rho(\alpha_n).$$

Theorem 3.17. *The Lyapunov exponent λ can be expressed as*

$$\lambda = \int_I \mu(f_\alpha) d\rho(\alpha)$$

Proof. We begin by fixing $z \in V$. Putting f_α in equation 3.8, evaluating at z, integrating both sides with respect to α, swapping the integral and the limit (because $|\mathcal{L}^n f_\alpha(z)| \leq \|f_\alpha\|_\infty$ for all $\alpha \in I$ and all $z \in V$), then using that for any sequence $x_n \to x$, $n^{-1}\sum_{m=0}^{n-1} x_m \to x$, we have

$$\begin{aligned} \int_I \mu(f_\alpha) d\rho(\alpha) &= \lim_{n\to\infty} \int_I \mathcal{L}^n f_\alpha(x) d\rho(\alpha) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \int_I \mathcal{L}^m f_\alpha(x) d\rho(\alpha) \\ &= \lim_{n\to\infty} \frac{1}{n} \int_I \sum_{m=0}^{n-1} \mathcal{L}^m f_\alpha(x) d\rho(\alpha). \end{aligned} \tag{3.9}$$

For each n we have

$$\begin{aligned}\sum_{m=0}^{n-1}\int_I \mathcal{L}^m f_\alpha(x)d\rho(\alpha) &= \sum_{m=1}^{n}\int_I\int_I\cdots\int_I f_{\alpha_m}(T_{\alpha_{m+1}\cdots\alpha_n}(x))d\rho(\alpha_1)\cdots d\rho(\alpha_n)\\ &= \int_I\int_I\cdots\int_I\sum_{m=1}^{n} f_{\alpha_m}(T_{\alpha_{m+1}\cdots\alpha_n}(x))d\rho(\alpha_1)\cdots d\rho(\alpha_n)\\ &= \int_I\cdots\int_I f_{\underline{\alpha}}(x)d\rho(\alpha_1)\cdots d\rho(\alpha_n).\end{aligned}$$

Hence combining this equation and equation 3.9 we get

$$\int_I \mu(f_\alpha)d\rho(\alpha) = \lim_{n\to\infty}\frac{1}{n}\int_I\cdots\int_I f_{\underline{\alpha}}(x)d\rho(\alpha_1)\cdots d\rho(\alpha_n).$$

Using lemma 3.4 we can write the Lyapunov exponent of the family of matrices $(A_\alpha)_{\alpha\in I}$ as

$$\lambda = \lim_{n\to\infty}\frac{1}{n}\int_I\cdots\int_I f_{\underline{\alpha}}(x_{\underline{\alpha}})d\rho(\alpha_1)\cdots d\rho(\alpha_n).$$

If $\int_I \mu(f_\alpha)d\rho(\alpha)\neq\lambda$ then there exists $\epsilon>0$ and for any $N\in\mathbb{N}$ there exists $n>N$ such that

$$\int_I\cdots\int_I |f_{\underline{\alpha}}(x_{\underline{\alpha}}) - f_{\underline{\alpha}}(x)|d\rho(\alpha_1)\cdots d\rho(\alpha_n) > n\epsilon,$$

which implies there exists $\underline{\alpha}\in I^n$ such that

$$|f_{\underline{\alpha}}(x_{\underline{\alpha}}) - f_{\underline{\alpha}}(x)| > n\epsilon,$$

but for large enough n this contradicts the bounded variation of $f_{\underline{\alpha}}$ given by lemma 3.12. □

For a fixed $\beta\in I$, we have a maximal simple eigenvalue of $\mathcal{L}_{0,\beta}$, and we now apply the perturbation theorem 2.2 to study how it changes for small t. Let $\epsilon>0$ be from the perturbation theorem such that we may define an analytic family of eigenvalues $\lambda_t^{(\beta)}$ for $\mathcal{L}_{\beta,t}$, $t\in B(0,\epsilon)\subset\mathbb{C}$, with $\lambda_0^{(\beta)}=1$.

Lemma 3.18. *We have*

$$\mu(f_\beta) = \frac{\partial \lambda_t^{(\beta)}}{\partial t}|_{t=0}.$$

Proof. Let $\lambda_t^{(\beta)}$ correspond to the eigenfunction $w_t^{(\beta)}$. We have that $\lambda_0^{(\beta)} = w_0^{(\beta)} = 1$ for all β. Differentiating the equation $\mathcal{L}_{\beta,t} w_t^{(\beta)} = \lambda_t^{(\beta)} w_t^{(\beta)}$ with respect to t gives us

$$\mathcal{L}_{\beta,t}\left(\frac{\partial w_t^{(\beta)}}{\partial t} + w_t^{(\beta)} f_\beta\right) = \frac{\partial \lambda_t^{(\beta)}}{\partial t} w_t^{(\beta)} + \lambda_t^{(\beta)} \frac{\partial w_t^{(\beta)}}{\partial t},$$

and at $t = 0$ this becomes,

$$\mathcal{L}\left(\frac{\partial w_t^{(\beta)}}{\partial t}|_{t=0} + f_\beta\right) = \frac{\partial \lambda_t^{(\beta)}}{\partial t}|_{t=0} + \frac{\partial w_t^{(\beta)}}{\partial t}|_{t=0}.$$

We now integrate both sides of this equation with respect to μ, and use the $\mathcal{L}$ invariance of μ to get

$$\int_U f_\beta d\mu = \frac{\partial \lambda_t^{(\beta)}}{\partial t}|_{t=0}.$$

□

We combine the previous two results to give the following corollary.

Corollary 3.19. *We can write the Lyapunov exponent as*

$$\lambda = \int_I \frac{\partial \lambda_t^{(\beta)}}{\partial t}|_{t=0} d\rho(\beta). \tag{3.10}$$

Since it is fortunate that this expression involves the measure ρ rather than μ, perhaps it is possible to use this corollary to numerically approximate the Lyapunov exponent, using the following ideas:

- Performing numerical integration for equation 3.10 requires the values of $\frac{\partial \lambda_t^{(\beta)}}{\partial t}|_{t=0}$ for $\beta \in J$ where J is some finite subset of I.
- We could approximate the transfer operators $\{\mathcal{L}_{t,\beta}\}_{\beta \in J}$ using finite matrices.

- We could then find the eigenvalues using a standard algorithm, e.g. the QR algorithm, and calculate the derivative using a simple finite difference rule.

The difficulty is approximating the transfer operators with finite matrices. We attempt to calculate the eigenvalue derivatives using dynamical determinants, which we do in the next section.

3.4 Dynamical Determinants

We wish to find an explicit formula for the trace of $\mathcal{L}^n_{\beta,t}$, so that we can use the theory in section 2.2. We start by fixing $\underline{\alpha} \in I^n$ and considering the simpler operator $\mathcal{L}_{\underline{\alpha}} : F(V) \to F(V)$ defined for $w_{\underline{\alpha}} \in F(V)$, by

$$\mathcal{L}_{\underline{\alpha}} f = f \circ T_{\underline{\alpha}} \cdot w_{\underline{\alpha}}.$$

Proposition 3.20. *As usual let $x_{\underline{\alpha}}$ denote the unique fixed point of $T_{\underline{\alpha}}$. Then*

$$\mathrm{Tr}\,(\mathcal{L}_{\underline{\alpha}}) = \frac{w(x_{\underline{\alpha}})}{\det(I - DT_{\underline{\alpha}}(x_{\underline{\alpha}}))}.$$

Proof. See Ruelle (47) or Bandtlow and Jenkinson (5). This is a general result, since $T_{\underline{\alpha}}$ is a holomorphic map with a unique fixed point and w is its weighting function. □

Next if we fix $t \in \mathbb{C}$ of small modulus, $\beta \in I$, and define

$$w_{\underline{\alpha}}(x) = \exp\left(t \sum_{m=1}^{n} f_\beta(T_{\alpha_m \alpha_{m+1} \cdots \alpha_{n-1} \alpha_n} x) \right)$$

then

$$\mathcal{L}^n_{\beta,t} f = \int_I \cdots \int_I \mathcal{L}_{\underline{\alpha}} f \, d\rho(\alpha_1) \cdots d\rho(\alpha_n).$$

Since the trace is continuous, we may use an approximation argument to give

$$\begin{aligned}\mathrm{Tr}\,(\mathcal{L}^n_{\beta,t}) &= \int_I \cdots \int_I \mathrm{Tr}\,\mathcal{L}_{\underline{\alpha}} d\rho(\alpha_1)\cdots d\rho(\alpha_n)\\ &= \int_I \cdots \int_I \frac{w_{\underline{\alpha}}(x_{\underline{\alpha}})}{\det(I - DT_{\underline{\alpha}}(x_{\underline{\alpha}}))} d\rho(\alpha_1)\cdots d\rho(\alpha_n)\end{aligned}$$

We can give a slightly easier expression for the denominator using the following lemma.

Lemma 3.21. *For any* $\underline{\alpha} \in I^n$, $n \in \mathbb{N}$, *we have*

$$\det(I - DT_{\underline{\alpha}}(x_{\underline{\alpha}})) = \prod_{i=1}^{d-1} \left(1 - \frac{\lambda_{i+1}}{\lambda_1}\right),$$

where $\lambda_1, \ldots, \lambda_d$ *are the eigenvalues of* $A_{\underline{\alpha}}$.

Proof. Following the proof of lemma 3.4, we may write in Jordan form $A_{\underline{\alpha}} = QJQ^{-1}$ with the first block of J being the maximal positive eigenvalue, and we can show

$$I - DT_{\underline{\alpha}}(x_{\underline{\alpha}}) = (DT_{Q^{-1}}(x_{\underline{\alpha}}))^{-1}(I - DT_J(T_{Q^{-1}}(x_{\underline{\alpha}})))DT_{Q^{-1}}(x_{\underline{\alpha}}),$$

and $T_{Q^{-1}}(x_{\underline{\alpha}}) = 0$, hence $\det(I - DT_{\underline{\alpha}}(X_{\underline{\alpha}})) = \det(I - DT_J(0))$. The eigenvalues of $DT_J(0)$ are λ_{i+1}/λ_1 where λ_i are the diagonal elements of J, for $j = 1, \ldots, d$. The result follows from this. □

In particular, for 2×2 matrices, we have

$$\det(I - DT_{\underline{\alpha}}(x_{\underline{\alpha}})) = 1 - \frac{\det A_{\underline{\alpha}}}{\lambda^2_{\underline{\alpha}}}.$$

Alternatively it might be possible to iterate $T_{\underline{\alpha}}$ a few times on any value $x \in V$ as noted in 3.8, and approximate the value of the denominator of the trace from this.

With all this in mind, for $z \in \mathbb{C}$ in a suitable neighbourhood, small $t \in \mathbb{C}$ and a fixed $\alpha \in I$ we define

$$\Delta_\beta(z,t) = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \int_I \cdots \int_I \frac{\exp\left(t \sum_{m=1}^{n} f_\beta(T_{\alpha_m \cdots \alpha_n} x_{\underline{\alpha}})\right)}{\det(I - DT_{\underline{\alpha}}(x_{\underline{\alpha}}))} d\rho(\alpha_1) \cdots d\rho(\alpha_n) \right).$$

Then the zeroes of this function are reciprocals of eigenvalues of $\mathcal{L}_{\beta,t}$. Let $z_t^{(\beta)}$ denote the zero corresponding to the maximal positive eigenvalue. So for small t,

$$\Delta_\beta(z_t^{(\beta)}, t) = 0,$$

and we can apply the implicit function theorem to this to get

$$\frac{\partial \Delta_\beta}{\partial z}(z_t^{(\beta)}, t) \frac{\partial z_t^{(\beta)}}{\partial t} + \frac{\partial \Delta_\beta}{\partial t}(z_t^{(\beta)}, t) = 0.$$

At $t = 0$, re-arranging, we have

$$\frac{\partial z_t^{(\beta)}}{\partial t}|_{t=0} = -\frac{\partial \Delta_\beta}{\partial t}(1,0) / \frac{\partial \Delta_\beta}{\partial z}(1,0).$$

Since

$$\frac{\partial \lambda_t^{(\beta)}}{\partial t} = -(z_t^{(\beta)})^{-2} \frac{\partial z_t^{(\beta)}}{\partial t},$$

we have the following as a result of corollary 3.19.

Corollary 3.22. *The Lyapunov exponent can be expressed as*

$$\lambda = \int_I \frac{\partial \Delta_\alpha}{\partial t}(1,0) \Big/ \frac{\partial \Delta_\alpha}{\partial z}(1,0) d\rho(\alpha)\ .$$

Unfortunately, the n integrals in the trace term makes it difficult to calculate numerically. The following ideas might be alternative approaches to calculating the continuous Lyapunov exponent:

- As previously mentioned, use lemma 3.19, and approximate the transfer operators using matrices.

- Approximate the family of matrices using matrix-valued step functions. This would give rise to a finite family of matrices, and an atomic probability measure, therefore we could calculate the Lyapunov exponent associated with this family using the discrete version. We would also require theory which tells us how to choose the step functions, but the hope is that the work already done in this chapter would provide an avenue for this.

3.5 The Discrete Case

As mentioned in the introduction, if we take $I = \{1, \ldots, m\}$ and $(p_1, \ldots, p_m)$ a probability vector, and define a probability measure $\rho = \sum_{i=1}^{m} p_i \delta_i$, then this reduces the setting to the discrete case. It is easy to show that in this situation we have

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \sum_{i_1,\ldots,i_n=1}^{k} p_{i_1} p_{i_2} \cdots p_{i_n} \log \|A_{i_1} A_{i_2} \cdots A_{i_n}\|.$$

Corollary 3.19 then becomes

$$\lambda = \sum_{j=1}^{m} p_j \, \frac{\partial \Delta_j}{\partial t}(1,0) \Big/ \frac{\partial \Delta_j}{\partial z}(1,0) \; ,$$

where the dynamical determinant $\Delta_j(z,t)$ is defined for $j \in I$, $z, t \in \mathbb{C}$ by

$$\Delta_j(z,t) = \exp\left(- \sum_{n=1}^{\infty} \frac{z^n}{n} \sum_{\underline{\alpha} \in I^n} \frac{\exp\left(t \sum_{m=1}^{n} f_\beta(T_{\alpha_m \cdots \alpha_n} x_{\underline{\alpha}})\right)}{\det(I - DT_{\underline{\alpha}}(x_{\underline{\alpha}}))} p_{\alpha_1} \cdots p_{\alpha_n} \right).$$

We can use this to calculate approximations of λ by expanding the determinant. This is very similar to the result of Pollicott in (43), except we work with linear fractional transformations instead of actions of positive matrices on the simplex. Similarly we can generalise the results here from Bernoulli measures to Markov measures.

Now we expand $\Delta_j(z,t)$ and take a truncation, as in equation 2.3 to define

$$\Delta_{j,N}(z,t) = 1 + \sum_{n=1}^{N} b_{j,n} z^n,$$

where

$$b_{j,n} = \sum_{\substack{n_1,\dots,n_m \in \mathbb{N} \\ n_1+\cdots n_m = n}} \frac{(-1)^m}{m!} \prod_{i=1}^{m} \frac{1}{n_i} \sum_{\underline{\alpha} \in I^{n_i}} \frac{\exp\left(t \sum_{m=1}^{n} f_\beta(T_{\alpha_m \cdots \alpha_n} x_{\underline{\alpha}})\right)}{\det(I - DT_{\underline{\alpha}}(x_{\underline{\alpha}}))} p_{\alpha_1} \cdots p_{\alpha_{n_i}}.$$

We use this to define a truncated version of the equation for λ given by corollary 3.19,

$$\lambda_N = \sum_{j=1}^{m} p_j \left. \frac{\partial \Delta_{j,N}}{\partial t}(1,0) \middle/ \frac{\partial \Delta_{j,N}}{\partial z}(1,0) \right. ,$$

so certainly $\lambda_N \to \lambda$ as $N \to \infty$. This shows that we can approximate the Lyapunov exponent in terms of fixed points and derivatives for the maps $T_{\underline{\alpha}}$. The analysis in the paper of Pollicott (43) which proves that the rate of convergence of λ_n is super-exponential also applies in this case. We will say more about the rate of convergence in the chapter on entropy rates of hidden Markov processes.

Example 3.23. The paper of Pollicott has an example where $I = \{1, 2\}$,

$$A_1 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix},$$

and $\underline{p} = (1/2, 1/2)$. We calculate the following approximations for λ. We implemented Mathematica code to calculate the following table.

n	λ_n
1	1.148425821572061634965821404607620388414268023042412
2	1.143086120634417773383184148181805684820891529663898
3	1.143311723765376730515258221104937606887902283334761
4	1.143311034849469294211018719581792364792031159094639
5	1.143311035102961529089783359627275487761860558182003
6	1.143311035102949245761627465541667784153770827818226
7	1.143311035102949245843251929555413172662505714965337
8	1.143311035102949245843251853655578309954402571112714
9	1.143311035102949245843251853655588299402733563872830
10	1.143311035102949245843251853655588299402546142483079

The computation in the paper of Pollicott is

$$\lambda \approx 1.1433110351029492458432518536555882994025,$$

which agrees with the calculation here.

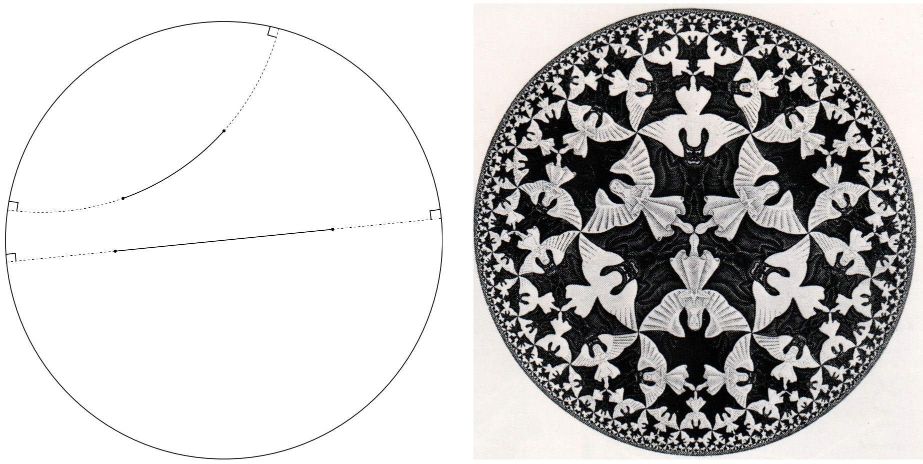
Chapter 4

Eigenfunctions of Laplacians on Surfaces of Constant Negative Curvature

Mark Kac posed the following famous question in an article (25) in 1966: "Can one hear the shape of a drum?" Since the shape determines the spectrum of the Laplacian and hence the sound, it is natural to look at the converse, and ask what we can determine about the geometry from the spectrum. Kac's question proved to be negative, as there are surfaces which have different geometry but have the same spectrum (for surfaces of constant negative curvature, there is a construction due to Sunada). However, the spectrum does contain a great deal of useful information.

We might hope to use dynamical determinants to say something about eigenfunctions of the Laplacian operator on surfaces of constant negative curvature. For these surfaces, the spectrum of the Laplacian is somewhat easier to compute, as eigenvalues correspond to zeros of the Selberg Zeta function or, equivalently, the determinant $\det(I - \mathcal{L}_s)$ of a suitable transfer operator $\mathcal{L}_s$.

We consider the case of a finite area surface K of constant negative curvature, of genus $g \geq 2$, and define the Laplacian operator on this. We aim to explore the



(a) (b)

Figure 4.1: (a) Geodesics on the Poincaré disk. (b) 'Angels and Devils' by M. C. Escher, 1941.

more challenging problem of analysing the eigenfunctions.

4.1 Background

4.1.1 Hyperbolic Geometry

Let $\mathbb{D}$ denote the unit disc $\{z \in \mathbb{C} : |z| < 1\}$. We obtain the Poincaré Disk Model for Hyperbolic space when we equip this with a Riemannian metric, given by

$$\langle u, v\rangle_z = \frac{4(u,v)}{(1-|z|^2)^2},$$

where $u, v \in T_z\mathbb{D}$ and $(\cdot,\cdot)$ is the usual Euclidean inner product. This gives rise to a metric in the usual way; given a curve $\gamma : [a,b] \to \mathbb{D}$ define $L(\gamma) = \int_a^b \langle \gamma'(t), \gamma'(t)\rangle_{\gamma(t)}^{1/2} dt$, then define $d(z,w) = \inf_\gamma L(\gamma)$ where the infimum is taken over all curves joining z and w. The metric has constant curvature -1. The geodesics in $\mathbb{D}$ correspond to circular Euclidean arcs that meet the boundary perpendicularly, and straight lines through the point $0 \in \mathbb{D}$, see figure 4.1(a).

Denote by $\mathcal{M}$ the subgroup of Mobiüs transformations given by all maps τ :

$\mathbb{D} \to \mathbb{D}$ of the form

$$\tau z = \frac{\alpha z + \beta}{\overline{\beta} z + \overline{\alpha}},$$

where $\alpha, \beta \in \mathbb{C}$ with $|\alpha|^2 - |\beta|^2 = 1$. It is readily verified that these are isometries for the Poincaré model, though they do not include reflections since $(z \mapsto \overline{z}) \notin \mathcal{M}$.

Let Γ be a discrete subgroup of $\mathcal{M}$ acting on $\mathbb{D}$, such that there are no elliptic elements (equivalent to there being no fixed points in the interior of $\mathbb{D}$) or parabolic elements (corresponding to having a single fixed point on the boundary), and let $K = \mathbb{D}/\Gamma$ be the equivalence classes of points related by actions of elements of Γ. The set K is given the quotient topology, i.e. the smallest one which makes continuous the projection $\pi : \mathbb{D} \to K$ given by $\pi x = \Gamma x$. This projection is locally a homeomorphism, since Γ contains no elliptical elements. The hyperbolic metric carries over, giving the surface K curvature of -1. It is a result from differential geometry that all complete, connected, compact two-dimensional Riemannian manifolds of curvature $\kappa = -1$ can be obtained in this way, as the quotient of such a subgroup of $\mathcal{M}$. Such surfaces are orientable with genus $g \geq 2$.

Definition 4.1. A closed subset $F \subset \mathbb{D}$ with interior F° is called a fundamental region of Γ if and only if:

1. $\tau_1 F^\circ \cap \tau_2 F^\circ = \varnothing$ if $\tau_1, \tau_2 \in \Gamma$ are distinct,

2. $\cup_{\tau \in \Gamma} \tau F = \mathbb{D}$, and

3. the boundary of F has zero two-dimensional Lebesgue measure.

Figure 4.1(b) demonstrates an image in a fundamental domain for some subgroup Γ, which has been repeated to tile the whole disc using isometries from Γ.

There are a number of ways of constructing fundamental regions, and when they exist, are far from unique. For example, τF is also a fundamental domain for $\tau \in \Gamma$. Often they are constructed using Dirichlet domains. We fix a point $u \in \mathbb{D}$ to be the origin, of the Dirichlet domain, and then define, for $\tau \in \Gamma$,

$$H_\tau(u) = \{z \in \mathbb{D} : d(z, u) < d(z, \tau u)\}$$

and define $H(u) = \cap_{\tau\in\Gamma} H_\tau(u)$ which is a fundamental domain with various properties, depending on Γ.

A fundamental region is known to exist for any Fuchsian group, see the book of Beardon (6), however if we restrict ourselves to surface groups, i.e. no elliptical elements and K is compact, then a fundamental region with further properties can be found. The following theorem is proved in a paper of Adler and Flatto (1):

Theorem 4.2. *Let K be a compact surface of genus $g \geq 2$, given by $\mathbb{D}/\Gamma$. There exists a bounded fundamental region F, which is a polygon whose boundary ∂F consists of $8g - 4$ geodesic segments, called edges. Label the edges $s_1, \ldots, s_{8g-4}$ in sequence with counter-clockwise orientation. Let s_i^{-1} denote s_i with opposite orientation. There are unique elements $\tau_i \in \Gamma$, $i = 1, \ldots, 8g - 4$ which map edges onto (reversed) edges, called side pairings, and a permutation σ of $1, \ldots, 8g - 4$, such that*

$$\tau_i(s_i) = s_{\sigma(i)}^{-1}.$$

Furthermore, F satisfies the extension condition, which states that the geodesics obtained from extending the edges remain completely in $\cup_{\tau\in\Gamma}\tau\partial F$, the Γ-orbit of ∂F.

It is also the case that the side pairings generate Γ, see the paper of Maskit (30).

Figure 4.2 shows an example of a fundamental domain with the extension condition. Note that there are very few fundamental domains that satisfy the extension condition. As observed by Pit (40), theorem 9.4.5 of (6) implies that for $p \in \mathbb{D}$ outside a set of measure zero, the Dirichlet domain of Γ based at p does not have the extension condition.

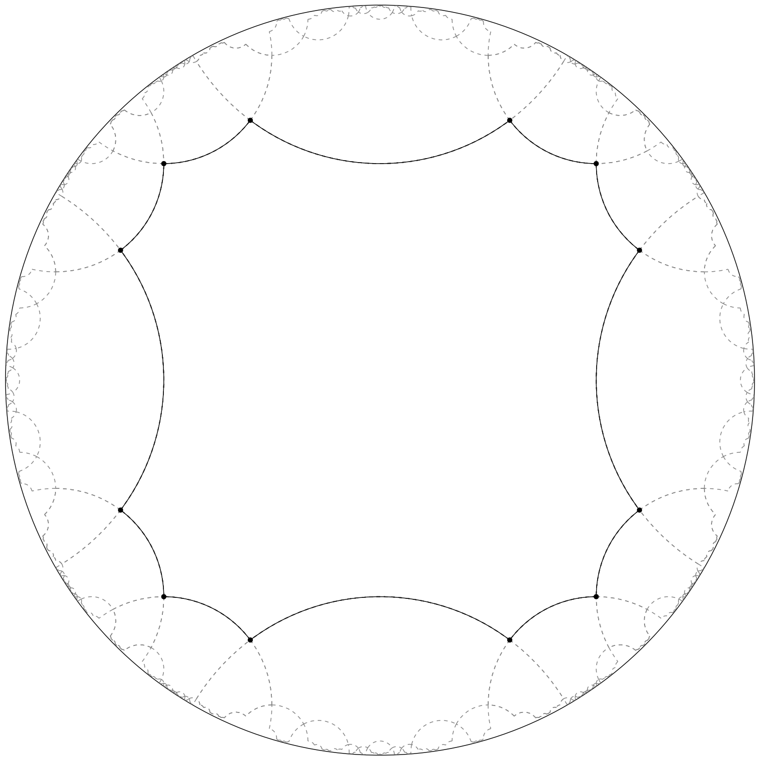


Figure 4.2: An example fundamental domain with the extension condition, and a small number of orbits.

4.1.2 The Bowen-Series Transformation

Assume the hypotheses of theorem 4.2 hold. Since the elements of Γ preserve the boundary, the fundamental polygon suggests the natural definition of a transformation on the boundary of $\mathbb{D}$, called the Bowen-Series transformation. Let $a_i, b_i \in \partial\mathbb{D}$ denote the start and endpoints of the geodesic ζ_i determined by s_i, with the same orientation as s_i. We assume the sides are listed counterclockwise, and that s_i is orientated in a counterclockwise so that moving along the geodesic from a_i to b_i is counterclockwise. By (a_i, b_i) we mean the open arc on $\partial\mathbb{D}$ given by moving counterclockwise from a_i to b_i, with closed and half-open arcs are defined similarly. See figure 4.3. The index i is taken modulus $8g - 4$. Let $\tau_i \in \Gamma$ be the transformation which pairs edge s_i its corresponding paired edge. Define the (right) Bowen-Series

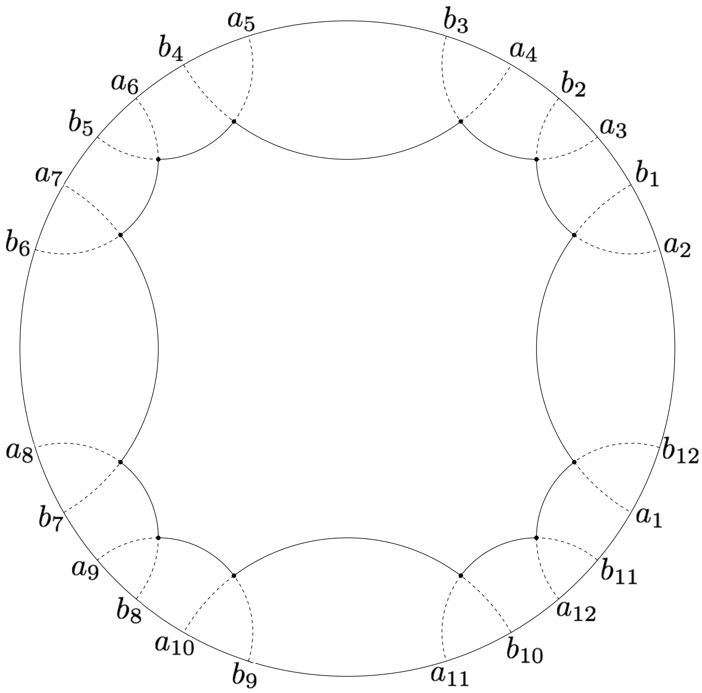


Figure 4.3: The arcs used to define the Bowen-Series transformation, for a surface of genus 2.

transformation $T : \partial\mathbb{D} \to \partial\mathbb{D}$ by

$$T(z) = \tau_i(z), \text{ whenever } z \in [a_i, a_{i+1}),$$

which is a definition for the whole of $\mathbb{D}$ since $[a_i, a_{i+1})$ is a partition. The refinement $\{I_k\}_{k=1}^{16g-8}$ is a Markov partition where

$$I_{2k-1} = [a_k, b_{k-1}), \quad I_{2k} = [b_{k-1}, a_{k+1}). \tag{4.1}$$

The Bowen-Series map was first studied by Bowen and Series in their classical paper (9). They study its connection to continued fractions, and show that the geodesic flow is ergodic, using that the Bowen-Series map is related to a cross section, and employing the following theorem. Adler terms this theorem the 'folklore theorem', on account of its provenance being difficult to determine. There is a discussion of this in Adler's afterword to the classical paper of Bowen (8). Appendix

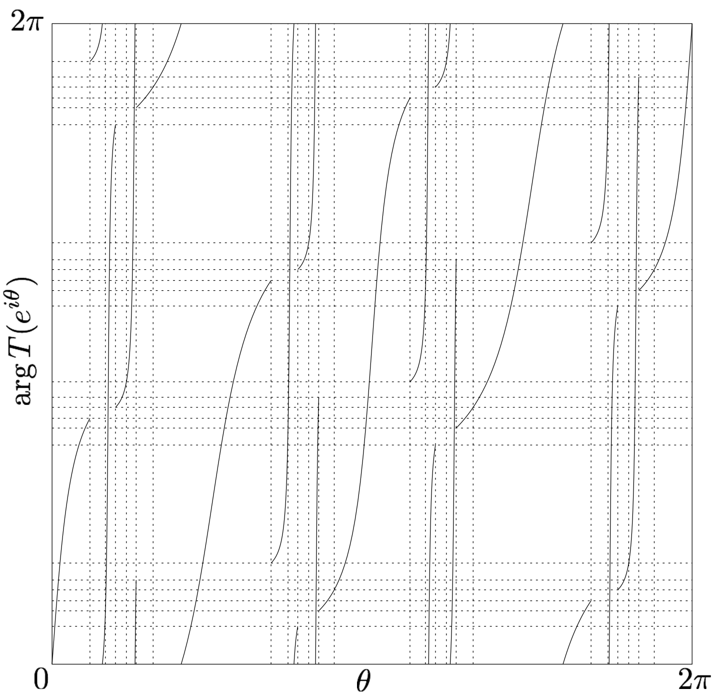


Figure 4.4: A graph of the Bowen-Series map for a surface of genus 2.

B of (1) contains a proof.

Theorem 4.3. *(Folklore Theorem) Let X be a 1-dimensional space, $\{I_i\}_{i=1}^N$ finite partition of X, and let $f : X \to X$ be such that, for each i,*

1. *$f|_{I_i}$ has a C^2 extension to $\overline{I_i}$,*
2. *$f|_{I_i}$ is invertible,*
3. *$f(\overline{I_i})$ is a union of I_j (the Markov property),*
4. *aperiodicity, i.e. for some p, $f^p(I_i) = X$, and*
5. *f is eventually expansive.*

Then there exists a measure μ that is f-invariant, ergodic, finite, and absolutely continuous, i.e. $d\mu = \rho dx$ with $1/D < \rho < D$ for some D.

The work of Helgason (16) relates eigenfunctions of the Laplacian to distributions on $\partial\mathbb{D}$. These turn out to be invariant for a transfer operator defined in terms of the Bowen-Series map.

Remark 4.4. The Bowen-Series map also codes information about closed geodesics. Given a fixed point $T^n x = x$, then we keep track of which side pairings T^n applies to x as we follow its orbit, and denote this as the map $\tau = \tau_{i_1}\tau_{i_2}\cdots\tau_{i_n}$. Then τ has two fixed points, $x, x' \in \partial\mathbb{D}$. The geodesic between them, $\gamma = [x, x']$, is the lift of a closed geodesic for the surface. Moreover, the length of γ is given by $\ell(\gamma) = \log|(T^n)'(x)|$.

4.1.3 The Hyperbolic Laplacian Operator

On a Riemannian manifold M, the Laplace-Beltrami operator $\Delta_M : C^\infty(M) \to C^\infty(M)$ is defined as the divergence of the gradient. This implies (for background, see the book of Morita (35)) a general local formula,

$$\Delta_M : f \mapsto \frac{1}{\sqrt{\overline{g}}}\sum_k \partial_k \left(\sum_i g^{ik}\sqrt{\overline{g}}\partial_i f\right),$$

where $g_{ij} = \left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right\rangle$, $\overline{g} = \det(g_{ij})_{i,j=1}$, and g^{ij} denotes the ij-entry of the inverse of the matrix given by $(g_{ij})_{i,j=1}$.

On the Poincaré disc, it reduces to

$$\Delta_{\mathbb{D}} : f(x+iy) \mapsto \frac{(1-x^2-y^2)^2}{4}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) f(x+iy).$$

The operator is invariant under automorphisms. For any $\tau \in \mathcal{M}$ and $f \in \mathcal{C}^\omega(\mathbb{D})$, we have $\Delta_{\mathbb{D}}(f\circ\tau) = (\Delta_{\mathbb{D}} f)\circ\tau$. Any other operator with this property is a polynomial of $\Delta_{\mathbb{D}}$.

We now define a function that gives us the eigenfunctions of $\Delta_{\mathbb{D}}$. For $z \in \mathbb{D}$, $b \in \partial\mathbb{D}$, let $\langle z, b\rangle \in \mathbb{R}$ denote the signed hyperbolic distance from 0 to the unique horocycle at b passing through z, positive if 0 is outside the horocycle, negative otherwise. In the Poincaré disc, horocycles are Euclidean circles in $\overline{\mathbb{D}}$ tangent to $\partial\mathbb{D}$. The distance the function measures is shown on figure 4.1.3. Note that $\langle z, b\rangle$ is constant when b is fixed and z moves on a fixed horocyle through b.

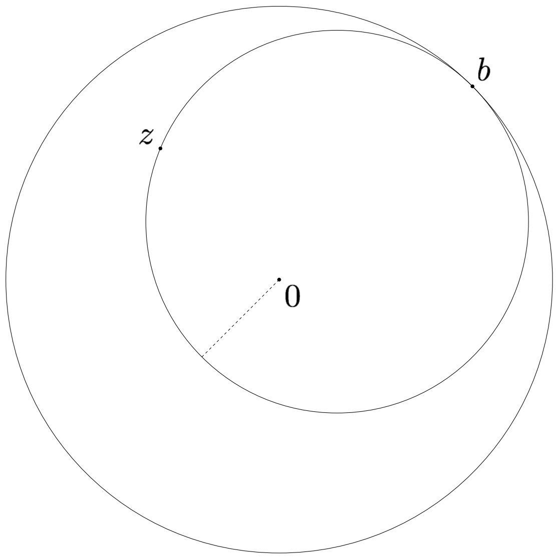


Figure 4.5: $\langle z, b\rangle$ is (in this case the negative of the) length of dotted line from 0 to horocycle through z and b.

Definition 4.5. For $b \in \partial\mathbb{D}$, and $\mu \in \mathbb{C}$ define $e_{\mu,b} : \mathbb{D} \to \mathbb{C}$ by

$$e_{\mu,b}(z) = e^{\mu\langle z,b\rangle}.$$

The following lemma tells us that this function is equal to the Poisson kernel when $\mu = 1$.

Lemma 4.6. *For $z \in \mathbb{D}$ and $b \in \partial\mathbb{D}$, we have that $e_{1,b}(z) = e^{\langle z,b\rangle} = \frac{1-|z|^2}{|b-z|^2}$.*

Proof. Let ξ denote the horocycle through b and z. Its centre is tb where $t \in (0,1)$, and the radius is $1-t$. Solving $|tb - z| = r$ gives $t = (1-|z|^2)(2 - 2(\mathfrak{Re}\, b\, \mathfrak{Re}\, z + \mathfrak{Im}\, b\, \mathfrak{Im}\, z))^{-1}$. Let w denote the point on ξ closest to zero. We have $|w| = 2t-1$. All that remains is to substitute the equation for t into $e^{d(0,w)} = \frac{1+|w|}{1-|w|}$. □

Lemma 4.7. *For $\tau \in \mathcal{M}$, $b \in \partial\mathbb{D}$, and $z \in \mathbb{D}$, we have:*

1. *$\langle \tau z, \tau b\rangle = \langle z, b\rangle + \langle \tau 0, \tau b\rangle$. (cf. (16), p. 83).*

2. *$|\tau'(b)| = \exp -\langle \tau 0, \tau b\rangle$. (cf. (16), p. 58).*

There now follows some results from Helgason (16), who uses the hyperbolic disc with curvature -4. We state the results here for -1, which is consistent with Pollicott (42) and Pit (40). This change just multiplies the metric by a constant.

Lemma 4.8. *Fix $b \in \partial\mathbb{D}$ and $\mu \in \mathbb{C}$. We have*

$$\Delta_{\mathbb{D}} e_{\mu,b} = \mu(\mu - 1)e_{\mu,b}.$$

Proof. Write $z = x + iy$ and $b = c + id$. The proof can be accomplished by simple calculations, since lemma 4.6 tells us $\langle z, b\rangle = \log f(x, y)$ where $f(x, y) = (1 - x^2 - y^2)((c - x)^2 + (d - y)^2)^{-1}$. We then have $\frac{\partial e_{\mu,b}}{\partial x} = \mu e_{\mu,b} \frac{1}{f} \frac{\partial f}{\partial x}$, and a similar derivative for y, which we can then differentiate again and substitute into the definition of $\Delta_{\mathbb{D}}$. □

Let $\mathcal{A}$ denote the space of analytic functions whose domain includes $\partial\mathbb{D}$. An *analytic distribution* ν is a continuous linear functional on $\mathcal{A}$, i.e. an element of the dual space $\mathcal{A}^\star$. Since this generalises measures, we use the notation $\int_{\partial\mathbb{D}} g d\nu = \nu(g)$ for $g \in \mathcal{A}$. The next result of Helgason gives us the relationship between eigenfunctions of the Laplacian and these analytic distributions. For $\lambda \in \mathbb{C}$, let $\epsilon_\lambda(\mathbb{D})$ denote the eigenspace $\epsilon_\lambda(\mathbb{D}) = \left\{ f \in C^\infty(\mathbb{D}) \mid \Delta_{\mathbb{D}} f = -\frac{\lambda^2+1}{4} f \right\}$. Then we have:

Theorem 4.9. *((16), Theorem 4.3, p. 60) The eigenfunctions of $\Delta_{\mathbb{D}}$ are the functions*

$$f(z) = \int_{\partial\mathbb{D}} e_{s,b}(z) d\nu(b),$$

where $s = \frac{i\lambda+1}{2}$, $\lambda \in \mathbb{C}$ and $\nu \in \mathcal{A}^\star$. Moreover, if $i\lambda \neq -1, -3, -5, \ldots$, then $\nu \mapsto f$ is a bijection of $\mathcal{A}^\star$ to $\epsilon_\lambda(\mathbb{D})$.

A natural question is to ask how this relates to eigenfunctions for the Laplacian on surfaces of constant negative curvature. If K is a surface given by $\mathbb{D}/\Gamma$, and $h_K \in C^\infty(K)$ is an eigenfunction for Δ_K, then h_K lifts to a function $h_{\mathbb{D}}$ on $\mathbb{D}$

given by $h_K(\Gamma z) = h_{\mathbb{D}}(z)$. The function $h_{\mathbb{D}}$ is an eigenfunction for $\Delta_{\mathbb{D}}$, and it is automorphic, i.e. $h_{\mathbb{D}}(\tau z) = h_{\mathbb{D}}(z)$ for all $z \in \mathbb{D}$ and $\tau \in \Gamma$. However, by theorem 4.9 and the first part of lemma 4.7, we can write

$$h_{\mathbb{D}}(\tau z) = \int_{\partial\mathbb{D}} e_{s,b}(\tau z) d\nu(b) = \int_{\partial\mathbb{D}} e_{s,b}(z) e^{s\langle \tau 0, \tau b\rangle} d\nu(\tau b)$$

which implies, by the second part of lemma 4.7,

$$\tau^\star \nu = e^{-s\langle \tau 0, \tau(\cdot)\rangle} \nu = |\tau'|^s \nu.$$

By $\tau^\star \nu$ we mean the usual pullback $\tau^\star \nu(f) = \nu(f \circ \tau)$. Thus linear functionals which satisfy

$$\tau^\star \nu = |\tau'|^s \nu \tag{4.2}$$

for all $\tau \in \Gamma$ give rise to eigenfunctions for the laplacian on the surface $K = \mathbb{D}/\Gamma$. This gives us a version of theorem 4.9.

Theorem 4.10. *Let $-\frac{\lambda^2+1}{4}$ be an eigenvalue of Δ_K, with $i\lambda \neq -1, -3, -5, \ldots$. Let $s = \frac{i\lambda+1}{2}$. Then there is a bijection between*

- *lifts to $\mathbb{D}$ of eigenfunctions of Δ_K whose eigenvalue is $-\frac{\lambda^2+1}{4} = s(s-1)$, and*
- *analytic distributions $\nu \in \mathcal{A}^\star$ satisfying $\tau^\star \nu = |\tau'|^s \nu$ for all $\tau \in \Gamma$.*

The distributions are invariant distributions for a transfer operator which is defined in terms of the Bowen-Series transformation, and this connection is explored in the next section.

By considering the extension of $\Delta_{\mathbb{D}}$ to the Hilbert space $L^2(\mathbb{D})$ and applying the theory of self-adjoint operators, it follows that the spectrum of $-\Delta_K$ consists of a countable sequence of eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \to \infty$. Each eigenvalue is counted according to its multiplicity.

Remark 4.11. For the special case $s = 1$, equation 4.2 is satisfied by the Patterson-Sullivan measure.

4.1.4 Transfer Operators

Following Pollicott (42) and Pit (40), we define a transfer operator for the Bowen-Series transformation. Let I be an interval of $\partial\mathbb{D}$ and let $\mathcal{C}^1(I)$ denote the space of complex-valued functions defined on I that are the restriction of a continuously differentiable function defined on an open neighbourhood of $\overline{I}$. The transfer operator $\mathcal{L}_s : E \to E$ associated with T for $s \in \mathbb{C}$ is defined by:

$$\mathcal{L}_s\phi(x) = \sum_{Ty=x} \frac{\phi(y)}{|T'(y)|^s}, \tag{4.3}$$

where

$$E = \{f : \partial\mathbb{D} \to \mathbb{C} : f|_{I_k} \in \mathcal{C}^1(I_k)\ \forall k\},$$

with $\{I_k\}_{k=1}^{16g-8}$ being the Markov partition defined by equation 4.1.

Pit extends earlier work of Pollicott with the following theorem:

Theorem 4.12 (Pit (40))**.** *For every $s \in \mathbb{C}$ such that $0 < \mathfrak{Re}(s) \leq 1$, there is an isomorphism between:*

- *The space of linear functionals ν that can be written as the weak derivative of a $\mathfrak{Re}(s)$-Hölder function and that satisfy, for all $\phi \in E$,*

$$\int_{\partial\mathbb{D}} \mathcal{L}_s\phi d\nu = \int_{\partial\mathbb{D}} \phi d\nu.$$

- *The space of smooth bounded Γ-invariant eigenfunctions of the hyperbolic Laplacian in $\mathbb{D}$ for the eigenvalue $s(1-s)$.*

It is given by $\nu \mapsto (z \mapsto \int_{\partial\mathbb{D}} e_{s,b}(z)d\nu(b))$.

In other words, if we find the right linear functional ν, i.e. $\mathcal{L}_s^\star \nu = \nu$, then we can recover an eigenfunction h using

$$h(z) = \int_{\partial\mathbb{D}} e_{s,b}(z) d\nu(b). \tag{4.4}$$

In the main part of this chapter, section 4.2, we study how to find these $\mathcal{L}_s$-invariant distributions, making use of the connection to a slightly different transfer operator. This will allow us to express the value of the eigenfunction at some $z \in \mathbb{D}$ as a rapidly converging series involving only fixed points of the Bowen-Series map.

4.1.5 Computing Eigenvalues

For purposes of computation involving the Laplacian, we need to have good approximations of the eigenvalues.

The Bowen-Series map $T : \partial\mathbb{D} \to \partial\mathbb{D}$ has a Markov partition $\{I_i\}_{i=1}^N$, $N = 16g-8$, given by equation 4.1. We can choose an open complex neighbourhood U_i of $\overline{I_i}$ such that if $T(I_i) \supset I_j$ then $T(U_i)^\circ \supset U_j$. Let $U = \coprod_{i=1}^N U_i = \cup_{i=1}^N(\{i\} \times U_i)$, the disjoint union of all the U_i's. $f \in F(U)$ means $f_i \in F(U_i)$ for all i where $f_i = f(i, \cdot)$ is referred to as a *component* of f. If $T(U_i) \supset U_j$ then the inverse branch is a function $T_{ij} : U_j \to U_i$ satisfying $T \circ T_{ij} = id_{U_j}$. Define a $N \times N$ matrix A by $A(i,j) = 1$ if $T(U_i) \supset U_j$ and 0 otherwise. Moreover, due to our subsequent choice of weight functions, we must follow remark 2.14 and take these domains to be open subsets of $\mathbb{C}^2$.

The set of admissible words of length n for A is

$$\Sigma_n = \{(i_1, i_2, \ldots, i_n) \,|\, A(i_j, i_{j+1}) = 1,\, j = 1, \ldots, n-1\}.$$

Definition 4.13. We define a transfer operator $\mathcal{M}_s : F(U) \to F(U)$ for $s \in \mathbb{C}$.

For $f \in F(U)$, $\mathcal{M}_s f$ is given by its components. So fixing $1 \le i \le N$ and $z \in U_i$,

$$(\mathcal{M}_s f)_i(z) = \sum_{j\,:\,A(j,i)=1} \frac{f_j(T_{ji}(z))}{|T'(T_{ji}(z))|^s} = \sum_{j\,:\,A(j,i)=1} f_j(T_{ji}(z))|T'_{ji}(z)|^s.$$

Note that the ith component of $\mathcal{M}_s$ is itself an operator $F(U) \to F(U_i)$. We denote it by $\mathcal{M}_{s,i}$.

The operator $\mathcal{M}_s$ is nuclear, because each component is a sum of nuclear composition operators, hence $\mathcal{M}_s$ is trace class. Calculating the trace of $\mathcal{M}_s^n$ involves the study of iterates of the partial inverses, it is convenient to introduce some notation. For $\underline{i} \in \Sigma_n$, define $T_{\underline{i}} = T_{i_1 i_2} \circ T_{i_2 i_3} \circ \cdots \circ T_{i_{n-1} i_n}$, and $w_{s,\underline{i}}(x) = |T'_{\underline{i}}(x)|^s$. We have $T_{\underline{i}} : U_{i_n} \to U_{i_1}$ and $w_{s,\underline{i}} \in F(U_{i_n})$. We define the composition operator $\mathcal{N}_{s,\underline{i}} : F(U_{i_1}) \to F(U_{i_n})$ by $\mathcal{N}_{s,\underline{i}} f(z) = f(T_{\underline{i}}(z)) w_{s,\underline{i}}(z)$.

For $\underline{i} \in \Sigma_n$, we define

$$U_{\underline{i}} = \cap_{j=1}^n T^{1-j} U_{i_j}.$$

Then in fact $T_{\underline{i}} : U_{i_n} \to U_{\underline{i}}$. For each element $\underline{i}$ of the set

$$\mathrm{Fix}(n) = \{\underline{i} \in \Sigma_{n+1} \,:\, i_{n+1} = i_1\}$$

we can show that there exists a unique element $x_{\underline{i}} \in U_{\underline{i}}$ such that $T_{\underline{i}} x_{\underline{i}} = x_{\underline{i}}$.

Lemma 4.14. *We have the following formula for the trace of $\mathcal{M}$:*

$$\operatorname{tr} \mathcal{M}_s^n = \sum_{\underline{i} \in \mathrm{Fix}(n)} \frac{|T'_{\underline{i}}(x_{\underline{i}})|^s}{|1 - T'_{\underline{i}}(x_{\underline{i}})|^2}.$$

Proof. We can verify that

$$\mathcal{M}_{s,j}^n f = (\mathcal{M}_s^n f)_j = \sum_{\substack{\underline{i} \in I_{n+1} \\ i_{n+1} = j}} \mathcal{N}_{s,\underline{i}} f_{i_1}.$$

Next fix $n \in \mathbb{N}$ and define $\mathcal{P}_{s,i}^n : F(U) \to F(U)$ by $(\mathcal{P}_{s,i}^n f)_i = \mathcal{M}_{s,i}^n f$ and $(\mathcal{P}_{s,i}^n f)_j = 0$

for $j \neq i$. This gives $\mathcal{M}_s^n = \sum_{i=1}^N \mathcal{P}_{s,i}^n$.

Next, fix k and assume $h \in F(U)$ is such that $\mathcal{P}_{s,k}^n h = \lambda h$. Hence $(\mathcal{P}_{s,k}^n h)_k = \mathcal{M}_{s,k}^n h = \lambda h_k$ and $h_i = 0$ for $i \neq k$. This gives

$$\lambda h_k = \sum_{\substack{\underline{i} \in I_{n+1} \\ i_{n+1}=k}} \mathcal{N}_{s,\underline{i}} h_{i_1} = \sum_{\substack{\underline{i} \in I_{n+1} \\ i_1 = i_{n+1}=k}} \mathcal{N}_{s,\underline{i}} h_k,$$

the second equality because h_{i_1} is only non-zero when $i_1 = k$. Conversely, if $h \in F(U_k)$ is such that $\sum_{\substack{\underline{i} \in \Sigma_{n+1} \\ i_1 = i_{n+1}=k}} \mathcal{N}_{s,\underline{i}} h = \lambda h$ then we can show that λ is an eigenvalue of $\mathcal{P}_{s,k}^n$. Putting all of this together gives

$$\operatorname{tr} \mathcal{M}_s^n = \sum_{i=1}^N \operatorname{tr} \mathcal{P}_{s,i}^n = \sum_{i=1}^N \operatorname{tr} \left(\sum_{\substack{\underline{i} \in \Sigma_{n+1} \\ i_1 = i_{n+1}=i}} \mathcal{N}_{s,\underline{i}} \right) = \sum_{\underline{i} \in \mathrm{Fix}(n)} \operatorname{tr} \mathcal{N}_{s,\underline{i}}.$$

For a fixed $\underline{i} \in \mathrm{Fix}(n)$ we have, from the work of Bandtlow and Jenkinson (5),

$$\operatorname{tr} \mathcal{N}_{s,\underline{i}} = \frac{w_{s,\underline{i}}(x_{\underline{i}})}{\det(1 - D\tilde{T}_{\underline{i}}(x_{\underline{i}}))},$$

where $\tilde{T}$ is the complexified transformation on $\mathbb{C}^2$. However, remark 2.14 gives us that this denominator is equal to $|1 - T_{\underline{i}}'(x_{\underline{i}})|^2$. This completes the proof. $\square$

The Bowen-Series map is defined piecewise in terms of Mobiüs maps $(\tau_i)_{i=1}^g$, $\tau_i : I_i \to \partial\mathbb{D}$, so if $A(i,j) = 1$ then $T_{ij} = \tau_i^{-1}|_{I_j}$. Thus $T_{\underline{i}} = \tau_{\underline{i}}^{-1}$ where $\tau_{\underline{i}} = \tau_{i_{n-1}} \circ \tau_{i_{n-2}} \circ \cdots \circ \tau_{i_1}$, and the fixed point $x_{\underline{i}}$ is the fixed point of $\tau_{\underline{i}}$. For the derivatives, we have that $T_{\underline{i}}'(x_{\underline{i}}) = 1/\tau_{\underline{i}}'(x_{\underline{i}})$. The weights are $w_{s,\underline{i}} = |\tau_{\underline{i}}'(x_{\underline{i}})|^{-s}$.

Proposition 4 in the paper of Pollicott (42) tells us that $\mathcal{M}_s$ has 1 as an eigenvalue if and only if $s(1-s)$ is an eigenvalue for Δ_K. Due to the operator being nuclear, and therefore the relation $\det(1 - \mathcal{M}_s) = \prod_{i=1}^\infty (1 - \lambda_i)$, we have that $\det(1 - \mathcal{M}_s) = 0$ if an only if 1 is an eigenvalue of $\mathcal{L}_s$. The following lemma is from (42).

Proposition 4.15. *The function $s \mapsto \det(1 - \mathcal{M}_s)$ is entire in s and its countably* *many zeros s_n are of the form $s_n = 1/2 + ir_n$, and correspond to the eigenvalues of the Laplacian as follows:*

$$\lambda_n = 1/4 + r_n^2$$

As usual, we can start with the formula for the determinant,

$$\det(1 - \mathcal{M}_s) = \exp\left(-\sum_{n=1}^{\infty} \frac{\operatorname{tr} \mathcal{M}_s^n}{n}\right),$$

and expand a power series, $\det(1 - \mathcal{M}_s) = 1 + \sum_{n=1}^{\infty} a_n(s)$, where

$$a_n(s) = \sum_{m=1}^{n} \sum_{\underline{i} \in \mathbb{N}^m : |\underline{i}| = n} \left(\frac{(-1)^m}{m!} \prod_{j=1}^{m} \frac{\operatorname{tr} \mathcal{M}_s^{i_j}}{i_j}\right).$$

This is detailed in equation 2.3. We can truncate this series, and find approximate positions of the zeros. Alternatively we can write $\mathcal{M}_s$ in the form $\sum_{n=0}^{\infty} \ell_n(\cdot) u_n$ and truncate this, to obtain an operator which maps polynomials to polynomials.

4.2 Computing Eigenfunctions

Now fix s so that $s(s-1)$ is an eigenvalue for Δ_K, where $K = \mathbb{D}/\Gamma$ and Γ is a surface group. Such values can be computed using section 4.1.5. The operator $\mathcal{M}_s$ being nuclear is therefore compact, so there exists a measure μ such that for all $f \in F(U)$, $\mathcal{M}_s^\star \mu = \mu$, or

$$\int_U \mathcal{M}_s f d\mu = \int_U f d\mu.$$

Integrating over a disjoint union means

$$\int_U f d\mu = \sum_{i=1}^{N} \int_{U_i} f_i d\mu_i \text{ where } \mu_i(V) = \mu(\{i\} \times V) \text{ for } V \subset U_i.$$

There is a natural embedding $\phi : F(U) \hookrightarrow E$. Given $f \in F(U)$ we define $(\phi f) : \partial\mathbb{D} \to \mathbb{C}$ by $(\phi f)|_{I_i} = f_i|_{I_i}$ for each i. Hence we may embed $\phi^\star : E^\star \hookrightarrow F(U)^\star$, using $(\phi^\star\nu)(f) = \nu(\phi f)$.

We would like to obtain eigenfunctions from μ using something similar to equation 4.4. Fix $z \in \mathbb{D}$. Define $f_{s,z} \in F(U)$ by $f_{s,z}(b) = e_{s,b}(z)$, using the formula for $e_{s,b}$ given by lemma 4.6 to provide the necessary extension. We first need a link between $\mathcal{L}_s$ and $\mathcal{M}_s$.

Lemma 4.16. *$\phi \circ \mathcal{M}_s = \mathcal{L}_s \circ \phi$.*

Proof. To see this fix $z \in I_j$ and $f \in F(U)$. Then $(\phi(\mathcal{M}_s f))|_{I_j}(z) = (\mathcal{M}_s f)_j|_{I_j}(z) = \sum_{y|Ty=z} \frac{(\phi f)(y)}{|T'(y)|^s} = \mathcal{L}_s(\phi f)(z)$. □

Theorem 4.17. *Fix $s \in \mathbb{C}$ such that $s(s-1)$ is a simple eigenvalue for Δ_K. Let $h \in C^\infty(\mathbb{D})$ be the associated eigenfunction. Then we have that*

$$h(z) = \mu(f_{s,z}).$$

Proof. Let ν be as in the statement of theorem 4.12. We show that μ is the embedding of ν, i.e. $\phi^\star\nu = \mu$. Using lemma 4.16 and the $\mathcal{L}_s$-invariance of ν, we have $(\phi^\star\nu)(\mathcal{M}_s f) = \nu(\phi(\mathcal{M}_s f)) = \nu(\mathcal{L}_s(\phi f)) = \nu(\phi f) = (\phi^\star\nu)(f)$ for all $f \in F(U)$, so $\phi^\star\nu$ is $\mathcal{M}_s$-invariant. Since μ is the unique $\mathcal{M}_s$-invariant distribution, then we must have that $\mu = \phi^\star\nu$. Finally, fixing $z \in \mathbb{D}$ and starting with equation 4.4, we have $h(z) = \int_{\partial\mathbb{D}} (\phi f_{s,z})(b) d\nu(b) = (\phi^\star\nu)(f_{s,z}) = \mu(f_{s,z})$. □

We then use ideas from the paper of Jenkinson and Pollicott (24) on integrating analytic functions. Fix z. We extend the transfer operator $\mathcal{M}_s$ with an additional weighting function to define a transfer operator $\mathcal{M}_{s,t,z} : F(U) \to F(U)$ as follows. Fix $t \in \mathbb{C}$, $g \in F(U)$. The ith component of $\mathcal{M}_{s,t,z}$ is given by:

$$(\mathcal{M}_{s,t,z} g)_i(x) = \sum_{j\,:\,A(j,i)=1} g_j(T_{ji}(x))|T'_{ji}(x)|^s e^{t f_{s,z}(T_{ji}(x))}.$$

Then $\mathcal{M}_{s,0,z}$ is the usual $\mathcal{M}_s$ operator.

If $\mathcal{M}_{s,t,z}\psi_{s,t,z} = \lambda_{s,t,z}\psi_{s,t,z}$, then noting that $\lambda_{s,0,z} = 1$, we can follow remark 2.13 to arrive at

$$\mu(f_{s,z}) = \left.\frac{\partial \lambda_{s,t,z}}{\partial t}\right|_{t=0} \int \psi_{s,0,z} d\mu,$$

however we can assume $\psi_{s,0,z}$ is appropriately normalised and integrates to 1.

We may adapt the proof of lemma 4.14 to get a trace formula for $\mathcal{M}_{s,t,z}$. We have

$$\operatorname{tr} \mathcal{M}_{s,t,z}^n = \sum_{\underline{i} \in \mathrm{Fix}(n)} \frac{|T_{\underline{i}}'(x_{\underline{i}})|^s \exp\left(t \sum_{m=1}^n f_{s,z}(T_{i_m i_{m+1} i_n} x_{\underline{i}})\right)}{|1 - T_{\underline{i}}'(x_{\underline{i}})|^2}.$$

We can now define the dynamical determinant $d_{s,z} : \mathbb{C}^2 \to \mathbb{C}$ for fixed $s \in \mathbb{C}, z \in \mathbb{D}$ by

$$d_{s,z}(w,t) = \exp\left(-\sum_{n=1}^{\infty} \frac{w^n}{n} \operatorname{tr} \mathcal{M}_{s,t,z}^n\right).$$

This function satisfies $d_{s,z}(\lambda_{s,t,z}^{-1}, t) = 0$ for small t. We apply the implicit function theorem to get

$$\left.\frac{\partial \lambda_{s,t,z}}{\partial t}\right|_{t=0} = \left.\frac{\partial d_{s,z}}{\partial t} \middle/ \frac{\partial d_{s,z}}{\partial w}\right|_{t=0,w=1}.$$

This gives us an expression for the eigenfunction h of the Laplacian.

Corollary 4.18. *If $\mathcal{M}_s$ has 1 as an eigenvalue, then*

$$\Delta_K h = s(s-1)h$$

where the eigenfunction $h \in C^\infty(\mathbb{D})$ is given by

$$h(z) = \left.\frac{\partial d_{s,z}}{\partial t} \middle/ \frac{\partial d_{s,z}}{\partial w}\right|_{t=0,w=1}.$$

Thus at least in principle, the eigenvalues and eigenfunction can be approximated using only periodic points of the Bowen-Series map T, by calculating a truncation of a series which converges super-exponentially.

4.3 Numerical Experiments

Following the paper of Pollicott and Rocha (45), we consider a surface K of genus 2 given in Fenchel-Nielsen co-ordinates $(\ell_1, \ell_2, \ell_3, 0, 0, 0)$, i.e. with zero twist angles. Nielsen gave a construction of a subgroup Γ and fundamental polygon which corresponds to K, and has the extension condition. An overview of the construction, represented in figure 4.6, is as follows.

1. The lengths ℓ_i, $i = 1, 2, 3$ correspond to lengths of closed geodesics on K, as seen in figure 4.6(a).

2. We next cut K into two identical pairs of pants at these geodesics, see figure 4.6(b).

3. Each pair of pants is cut along three more geodesics, labelled on figure 4.6(c), to get two right-angled hexagons, one of which is shown in figure 4.6(d).

4. Each of the four hexagons is copied into $\mathbb{D}$, to get a 12-sided fundamental polygon, as seen in figure 4.6(f). The edges are identified to correspond to the cuts we have made.

Because the lengths ℓ_i give us three sides of a right-angled hexagon, we can use an equation from Beardon (6) to obtain the other three sides. We define ℓ_i', $i = 1, 2, 3$ by

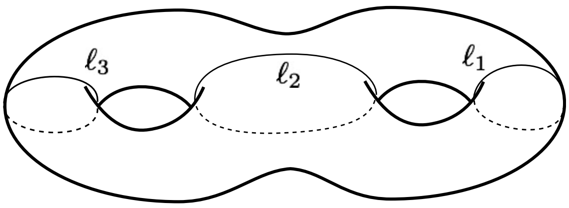
$$\cosh(\ell_1'/2) = \frac{\cosh(\ell_1/2) + \cosh(\ell_2/2)\cosh(\ell_3/2)}{\sinh(\ell_2/2)\sinh(\ell_3/2)},$$

$$\cosh(\ell_2'/2) = \frac{\cosh(\ell_2/2) + \cosh(\ell_1/2)\cosh(\ell_3/2)}{\sinh(\ell_1/2)\sinh(\ell_3/2)},$$

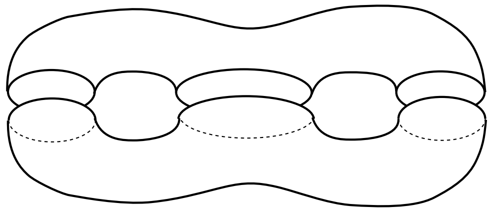
$$\cosh(\ell_3'/2) = \frac{\cosh(\ell_3/2) + \cosh(\ell_1/2)\cosh(\ell_2/2)}{\sinh(\ell_1/2)\sinh(\ell_2/2)}.$$

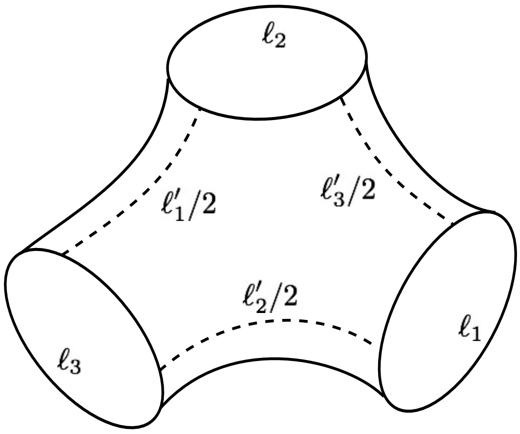
Remark 4.19. This construction can be generalised to surfaces of genus $g \geq 2$, by decomposing the surface into more pairs of (non-identical) pants.

Computationally, we can proceed as follows.

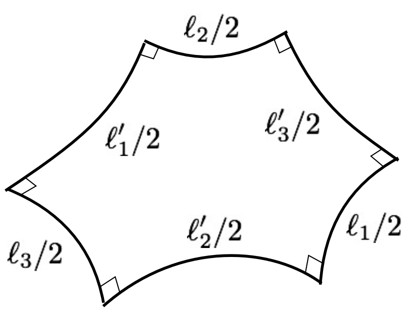


(a) The surface K and the three geodesics ℓ_i.

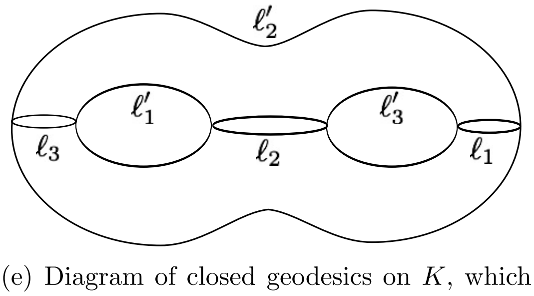

(b) K split into two identical pairs of pants.



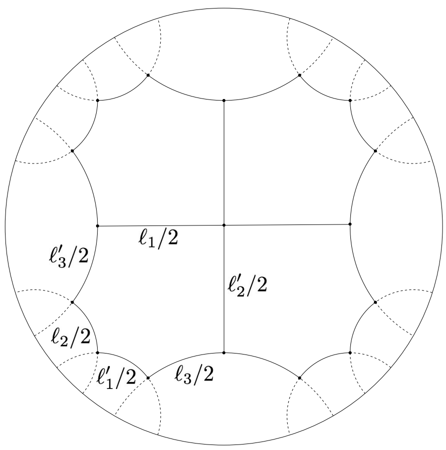
(c) One pair of pants can be split into two identical right-angled hexagons.



(d) One right-angled hexagon with lengths labelled.



(e) Diagram of closed geodesics on K, which cut to give the fundamental polygon.



(f) Four copies of the hexagon in $\mathbb{D}$, with edges labelled.

Figure 4.6: The construction of a fundamental domain for K.

1. We construct the fundamental polygon by analogy with the Turtle robot[1]. We can obtain the first hexagon by starting at 0, and moving in the direction -1 a hyperbolic distance of $\ell_1/2$ (which traces a geodesic), then turning clockwise $\pi/2$, moving hyperbolic distance $\ell'_3/2$, and so on. We can then reflect the hexagon by the imaginary axis to get the second hexagon. Finally we can reflect the first two by the real axis to get the second two. We then have all 12 vertices which make up the fundamental polygon. We can then calculate the side pairings by calculating the general isometry which takes one pair of vertices to another pair of vertices.

2. Use the fundamental polygon to calculate the endpoints of the arcs on $\partial\mathbb{D}$ which give us the Markov partition for the Bowen-Series map T, and store the element $\tau \in \Gamma$ for each interval.

3. Calculate the transition matrix A which gives us the symbolic dynamics for T. This can be achieved by taking the images of the end points of the intervals on $\partial\mathbb{D}$.

4. Given $\underline{i} \in \mathrm{Fix}(n)$, we need to calculate $x_{\underline{i}}$. Calculate $\tau_{\underline{i}} \in \Gamma$ using matrix multiplication, and solve the equation

$$z = \tau_{\underline{i}} z = \frac{\alpha z + \beta}{\overline{\beta} z + \overline{\alpha}},$$

which gives two solutions,

$$z_{\pm} = \frac{(\alpha - \overline{\alpha}) \pm \sqrt{(\overline{\alpha} - \alpha)^2 + 4|\beta|^2}}{2\overline{\beta}},$$

and $x_{\underline{i}}$ is the one with $|\tau'_{\underline{i}}(\cdot)| > 1$. This saves having to keep track of interval endpoints in order to find the fixed point.

[1]See http://en.wikipedia.org/wiki/Turtle_(robot).

We can therefore try to compute expansions of the determinants for $\mathcal{M}_s$ and $\mathcal{M}_{s,t}$. Unfortunately, we found that with 24 elements in the Markov partition, the number of periodic points grows too quickly for us to be able to calculate more than the first 5 terms of the power series, and 5 terms did not give us anything which suggested convergence.

The current implementation is in Python code using the SymPy symbolic maths package, and runs on one CPU core. We found this to be faster than using Mathematica. The problem is an inherently parallel one, and with a more suitable implementation could be run on a cluster of computers, and this would be able to calculate a few more terms.

Another potential avenue is to exploit any symmetry in the fundamental polygon. At least in this example, we enumerate many geodesics with the same lengths many times. A clever way of accounting for this might have the effect of reducing the rate at which the number of distinct periodic points grows. Figure 4.7 which shows all prime closed geodesics of period 3 for a particular surface certainly shows symmetry in the real and imaginary axes.

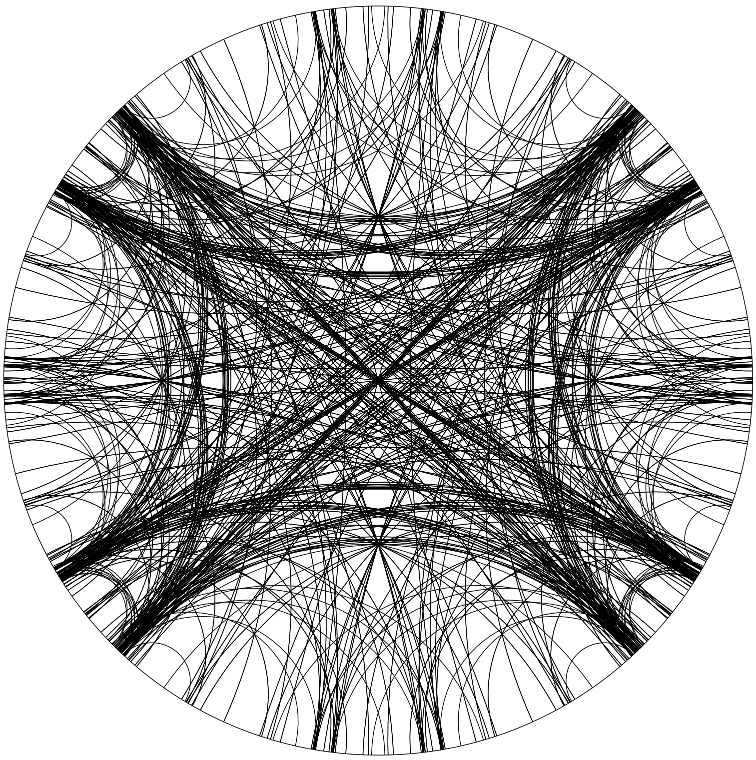


Figure 4.7: All closed prime geodesics of period 3 for a surface given by $\ell_1 = \ell_2 = \ell_3 = 2.63$.

Chapter 5

Approximations of Mahler Measures

5.1 Introduction

The Mahler measure $m(P)$ (or height) of a complex polynomial $P(z)$ is a numerical value which plays an important role in number theory, dynamical systems and geometry. It was introduced by Kurt Mahler as a device to provide a simple proof of Gelfond's inequality for the product of polynomials in many variables. However it has turned out to have much more versatile applications.

More generally, it occured in Lehmer's investigation of certain cyclotomic functions and led him to make what is now known as Lehmer's conjecture, about which polynomial has the least Mahler measure. It is conjectured to be $x^{10} + x^9 - x^7 - x^6 - x^5 - x^4 - x^3 + x + 1$, but in spite of considerable work by many authors on this conjecture, it has not yet been proved. In another direction, connections have been found between the Mahler measure of certain two variable polynomials and invariants of hyperbolic 3-manifolds such as the volume, via the dilogarithm and work of Milnor, Zagier and others. Finally, the Mahler measure of an integer polynomial in k variables gives the topological entropy of a $\mathbb{Z}^k$-dynamical system (for $k \geq 2$)

canonically associated to the polynomial in the work of Lind, Schmidt and Ward. These can be viewed as generalisations of Yuzvinskii's formula for automorphisms of solenoids in the case $k = 1$. However, unlike the case of $\mathbb{Z}$-actions these entropies need not be the logarithms of algebraic integers. The book (11) of Everest and Ward is a good reference for the Mahler measure and its applications.

There is no closed form for the Mahler measure and so for any given polynomial we can only expect to express it in terms of an infinite series, or something reflecting this property. There are different approaches to the Mahler measure, including relating it to zeta functions, as in the work of Smyth. Here we want to explore using the theory in the paper of Jenkinson and Pollicott (24), by rewriting the Mahler measure as the integral of an analytic function. We prove the following theorem for certain polynomials of one variable, but the hope is that it can be generalised.

Theorem 5.1. *We can write the logarithmic Mahler measure $m(P)$, of a polynomial P of one variable, as an infinite series*

$$m(P) = \sum_{n=1}^{\infty} a_n$$

where

1. *a_n is explicitly given in terms of the values of the polynomial P at the points $\{\frac{k}{2^n-1} : k = 0, 1, \cdots, 2^n - 1\}$;*
2. *There exists $0 < \theta < 1$ and $C > 0$ such that $|a_n| \leq C\theta^{n^2}$*

The use of infinite series as explicit expressions for the Mahler measure of polynomials is not new, occurring for example in the work of Smyth and others. In that case, it is related to expressions in terms of L-functions.

It is interesting that we use one dynamical method to evaluate a number theoretical value, given that one of the more recent applications is to computing the entropy of other dynamical systems.

5.2 Definition and Examples

We first define the logarithmic Mahler measure for a polynomial in one variable, and then extend this to the case of polynomials of several variables.

5.2.1 Polynomials in One Variable

We begin with the definition of the Mahler measure $M(P)$ for a complex valued polynomial $P(z)$. This is defined using the zeros and leading coefficient of the polynomial.

Definition 5.2. If $P(z) = a(z - \alpha_1) \cdots (z - \alpha_n)$ then

$$M(P) = |a| \prod_{|\alpha_i| \geq 1} |\alpha_i| \tag{5.1}$$

where the product is over zeros of modulus at least one. We also define the logarithmic Mahler measure by

$$m(P) = \log M(P).$$

From equation 5.1 it is easy to see that a cyclotomic polynomial has Mahler measure precisely 1, and that for the linear polynomial we have $M(az + b) = \max\{|a|, |b|\}$.

Remark 5.3. We can associate to a polynomial

$$f(x) = a_m x^m + \cdots + a_1 x + a_0$$

the matrix

$$A = \begin{pmatrix} -\frac{a_{m-1}}{a_m} & -\frac{a_{m-2}}{a_m} & \frac{a_{m-3}}{a_m} & \cdots & -\frac{a_1}{a_m} & -\frac{a_0}{a_m} \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

The eigenvalues of A are the zeros of the polynomial $f(x)$. Providing each zero $f(x) = 0$ has $|x| \geq 1$, we can write $M(P) = a_m \det A$, making it particularly easy to estimate directly.

There is another equivalent definition which is useful in making explicit numerical estimates. Using Jensen's theorem from complex analysis we can rewrite the defintion in the following form incorporating a definite integral:

$$M(P) = \exp\left(\frac{1}{2\pi}\int_0^{2\pi} \log|P(e^{i\theta})|d\theta\right). \tag{5.2}$$

The value of the Mahler measure can be explicitly computed using equation 5.2 in other simple examples.

Example 5.4. For the famous example of Lehmer's polynomial

$$P(x) = x^{10} + x^9 - x^7 - x^6 - x^5 - x^4 - x^3 + x + 1$$

the value of the logarithmic polynomial can be computed to be $m(P) = 1.176280818\ldots$. This is conjectured to be the smallest value for non-cyclotomic polynomials.

There are partial results, due to Smyth, which consider the restriction of the conjecture to a smaller class of polynomials. We call a polynomial $P(x)$ of degree d *reciprocal* if $x^d P(1/x) = P(x)$. Smyth (57) showed that if P is non-reciprocal then the smallest value is attained, and is equal to $m(z^3 - z - 1) = \log(1.324\ldots) = 0.281\ldots$.

5.2.2 Polynomials in Two Variables

Of more practical interest is the Mahler measure for polynomials of two variables. Moreover, it is more convenient to generalize the integral definition for the Mahler measure when considering functions of two (or more) complex variables.

Definition 5.5. For a polynomial of two variables $P(z_1, z_2)$ we can define the Mahler measure by

$$M(P) = \exp\left(\frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi}\log|P(e^{i\theta_1}, e^{i\theta_2})|d\theta_1 d\theta_2\right).$$

Example 5.6. When $P(z_1, z_2) = 1 + z_1 + z_2 - z_1 z_2$ then one can explicitly compute $m(P) = 1.7916228\ldots$.

There are explicit series for certain Mahler measures of particular polynomials, originally due to Smyth in 1981, which allow the Mahler measure to be written in terms of zeta functions and L-functions.

Example 5.7 (Smyth (58))**.** When $P(z_1, z_2) = 1 + z_1 + z_2$ we can write

$$m(1 + z_1 + z_2) = \frac{3\sqrt{3}}{4\pi}L(2, \chi_3)$$

where $L(s, \chi_3) = \sum_{n=1}^{\infty}\chi_3(n)n^{-s}$ is the L-function and thus

$$L(2, \chi_3) = 1 - \frac{1}{4} + \frac{1}{16} - \frac{1}{25} + \cdots$$

where

$$\chi_3(n) = \begin{cases} 0 & \text{if } n = 0 \bmod 3 \\ 1 & \text{if } n = 1 \bmod 3 \\ -1 & \text{if } n = 2 \bmod 3 \end{cases}$$

The characterisation of Mahler measure in terms of zeta-functions and L-functions extends to other examples.

Example 5.8. Following Smyth (58) we can also write

$$m(1 + z_1 + z_2 + z_3) = \frac{7}{2\pi^2}\zeta(3),$$

where $\zeta(s)$ is the Riemann zeta function. This identity comes from integrating series expansions term by term. However, it is difficult to numerically compute.

5.3 Modifying the Integrals

We would like to rewrite the integrals giving the Mahler measures in specific cases so that they are integrals of analytic functions defined in a neighbourhood of bounded intervals in the real line. This is an essential requirement of our approach.

Traditionally, the formulae for the logarithmic Mahler measure comes from expanding the integrands as a power series prior to integration. In particular, one can take the Taylor series of $\sin(\cdot)$ and $\cos(\cdot)$.

However we instead want to rewrite the definition in terms of an integral of an analytic function, perhaps over a different interval. This is illustrated by the following simple lemma, due to Smyth (58).

Lemma 5.9. *We can write*

$$m(1 + z_1 + z_2) = \int_{1/6}^{5/6} \log(2\sin(\pi y))dy.$$

Proof. Using Jensen's formula,

$$\int_0^{2\pi} \log|\alpha + \beta e^{it}|dt = 2\pi \log(\max\{|\alpha|, |\beta|\}).$$

We can write

$$\begin{aligned} m(1+z_1+z_2) &= \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi}\log|1+e^{i\theta_1}+e^{i\theta_2}|d\theta_1 d\theta_2 \\ &= \frac{1}{2\pi}\int_0^{2\pi}\log(\max\{|1+e^{i\theta_1}|,1\})d\theta_1 \\ &= \frac{1}{2\pi}\int_{-2\pi/3}^{2\pi/3}\log|1+e^{i\theta_1}|d\theta_1 \end{aligned}$$

since $|1+e^{i\theta_1}| \leq 1$ for $\theta_1 \in [2\pi/3, 4\pi/3]$. Finally, $|1+e^{i\theta_1}| = \sqrt{2+2\cos(\theta_1)}$, to which we apply a double angle formula for cos. □

The usual approach at this stage would be to expand

$$\log|1+e^{i\theta}| = \mathfrak{Re}\left(\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}e^{in\theta}\right)$$

and integrate term by term. The resulting series can then be interpreted as a suitable zeta function or L-function.

However, since $\sin(\pi y)$ is nonzero and analytic on $[1/6, 5/6]$ we see that $\log(2\sin(\pi y))$ is analytic in a neighbourhood U of the interval $[1/6, 5/6]$ which is a necessary ingredient in our approach. So as we will see later, we can express the integrand instead in terms of a summation involving periodic points for the doubling map.

In some cases the integral can be rewritten so that the singularity in the integrand can be removed and dealt with as a separate term, rather than eliminated completely. This is illustrated in the following lemma, again due to Smyth (58).

Lemma 5.10. *We can write*

$$m(1+z_1+z_2+z_3) = \frac{2}{\pi^2}\int_0^{\pi}\theta\log(2\sin(\theta/2))d\theta.$$

Proof. Using Jensen's formula again, we can write

$$\begin{aligned}
m(1+z_1+z_2+z_3) &= \frac{1}{(2\pi)^3}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\log|1+e^{i\theta_1}+e^{i\theta_2}+e^{i\theta_3}|d\theta_1 d\theta_2 d\theta_3 \\
&= \frac{1}{(2\pi)^3}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\log|(1+e^{i\theta_1})+e^{i\theta_3}(1+e^{i(\theta_2-\theta_3)})|d\theta_1 d\theta_2 d\theta_3 \\
&= \frac{1}{\pi^2}\int_0^{\pi}\int_0^{\pi}\log(\max\{|1+e^{iu}|,|1+e^{it}|\})dudt \\
&= \frac{2}{\pi^2}\int_0^{\pi}\int_t^{\pi}\log|1+e^{iu}|dudt \\
&= -\frac{2}{\pi^2}\int_0^{\pi} t\log|1+e^{it}|dt \\
&= \frac{2}{\pi^2}\int_0^{\pi}\theta\log\left(2\sin\frac{\theta}{2}\right)d\theta,
\end{aligned}$$

where we have used the double angle formula as before. □

As before the traditional approach would be to expand

$$\log|1+e^{i\theta}| = \mathfrak{Re}\left(\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}e^{in\theta}\right),$$

and integrate term by term. The resulting series can again then be interpreted as a suitable zeta function or L-function.

However, we see that we can write

$$\theta\log\left(2\sin\left(\frac{\theta}{2}\right)\right) = \theta\left(\log\left(\frac{2}{\theta}\sin\left(\frac{\theta}{2}\right)\right)+\log(\theta)\right)$$

and the first term on the RHS is analytic since we can expand it as a series about $\theta = 0$ to write:

$$\frac{2}{\theta}\sin\left(\frac{\theta}{2}\right) = 1+\sum_{n=1}^{\infty}\frac{(-1)^n}{(n+1)!}\left(\frac{\theta}{2}\right)^n.$$

In particular, the integrand $F:[0,\pi]\to\mathbb{R}$ defined by $F(\theta) = \theta\log\left(\frac{2}{\theta}\sin\frac{\theta}{2}\right)$ is analytic in a neighbourhood on $[0,\pi]$ allowing it to be efficiently integrated.

On the other hand the integral $\int_0^{\pi}\theta\log\theta d\theta$ can be evaluated as a standard integral (taking the value $\frac{1}{4}\pi^2(2\log\pi - 1)$). Thus the value of the Mahler measure can be written as the sum of an explicit special value (coming from the expression involving the singularity) and an explicit series (for F) which we can effectively evaluate.

5.4 Transfer Operators and Determinants

We now use the theory from the paper of Jenkinson and Pollicott (24) to prove theorem 5.1 in the case where we can re-write the Mahler measure to have an analytic integrand. Consider the contractions on the unit interval $T_0, T_1 : [0,1] \to [0,1]$ defined by

$$T_0(x) = \frac{x}{2} \quad \text{and} \quad T_1(x) = \frac{x+1}{2}.$$

There is no particular reason to take these contractions, except for the simplicity. We could expand in other bases, or even take non-linear contractions, but since this complicates computation and has no clear advantage, we prefer to use base 2.

The fixed points of $T_{i_1} \circ \cdots \circ T_{i_n}$ are of the form

$$\frac{k}{2^n - 1} \quad \text{where } 0 \leq k \leq 2^n - 1,$$

i.e., the periodic points for the doubling map.

Let $g : [0,1] \to \mathbb{R}$ be an analytic function (that is, one which extends analytically to a complex neighbourhood U of $[0,1]$). We can define expressions for $s \in \mathbb{C}$, $n \geq 1$ defined by

$$Z_n(s) = \frac{1}{2^n - 1} \sum_{k=0}^{2^n - 1} \exp\left(s \sum_{m=0}^{n-1} g\left(\left\{ \frac{2^m k}{2^n - 1} \right\} \right) \right),$$

where $\{\cdot\}$ denotes the fractional part of a number. We can then define the determinant $d : \mathbb{C}^2 \to \mathbb{C}$ given by

$$d(z,s) = \exp\left(-\sum_{n=1}^{\infty} \frac{Z_n(s)}{n} z^n \right)$$

Finally, we can expand this as a power series in z, as in equation 2.3.

Example 5.11. We can write

$$\begin{aligned} d(z,s) &= \exp\left(-zZ_1(s) - z^2\frac{Z_2(s)}{2} - z^3\frac{Z_3(s)}{3} + \dots\right) \\ &= 1 - \left(-zZ_1(s) - z^2\frac{Z_2(s)}{2} - z^3\frac{Z_3(s)}{3} + \dots\right) \\ &\quad + \frac{1}{2}\left(-zZ_1(s) - z^2\frac{Z_2(s)}{2} - z^3\frac{Z_3(s)}{3} + \dots\right)^2 + \cdots \\ &= 1 - z\underbrace{Z_1(s)}_{a_1(s)} + z^2\underbrace{\left(\frac{1}{2}(Z_1(s))^2 - \frac{1}{2}Z_2(s)\right)}_{=a_2(s)} + \cdots \end{aligned}$$

Theorem 5.12. *The function d is analytic for all $z, s \in \mathbb{C}$. In particular, we can write*

$$d(z,s) = 1 + \sum_{n=1}^{N} a_n(s)z^n + \mathcal{O}\left(2^{-N^2}\right)$$

Proof. See the paper of Jenkinson and Pollicott (24). □

We can also identify the following:

Lemma 5.13. *We have that*

$$\int g(x)dx = \frac{\partial d}{\partial s}(1,0)\left(\frac{\partial d}{\partial z}(1,0)\right)^{-1}.$$

Proof. This again comes from (24) and is an example of the method used in remark 2.13. Alternatively, note that $z(s) = e^{P(sg)}/2$ is the implicit solution to $d(z(s),s) = 0$. Then we can use that $P'(0) = \int g(x)dx$, and then apply the implicit function theorem. □

This allows us to get an explicit solution for the integral and a method of approximation, by truncating the determinant after a finite number of terms. More precisely, we can write

$$\int g(x)dx = \frac{\sum_{n=1}^{N} a_n'(0)}{\sum_{n=1}^{N} n a_n(0)} + \mathcal{O}\left(\theta^{N^2}\right) \tag{5.3}$$

for $N \geq 1$.

5.5 Numerical Evaluation

We can now apply equation 5.3 to calculate approximations to the integrals we previously modified to remove the singularities.

Example 5.14. We can evaluate the integral from lemma 5.9,

$$\int_{1/6}^{5/6} \log(2\sin(\pi y))dy$$

by first changing variables by $x = \frac{3}{2}(y - \frac{1}{6})$, to get

$$\frac{2}{3}\int_0^1 \log\left(2\sin\left(\frac{2\pi}{3}x + \frac{1}{6}\right)\right) dx.$$

This can then be evaluated using the periodic point expansion and equation 5.3, and we give of this equation truncated after N terms in table 5.1.

Example 5.15. We can evaluate the integral from lemma 5.10,

$$\frac{2}{\pi^2}\int_0^{\pi} \theta \log\left(\frac{2}{\theta}\sin\left(\frac{\theta}{2}\right)\right) d\theta,$$

by first changing variables by $x = \theta/\pi$, to get

$$2\int_0^1 x\log\left(\frac{2}{\pi x}\sin\left(\frac{\pi x}{2}\right)\right) = 3\underbrace{\int_0^1 x\log x dx}_{=-1/4} + 2\int_0^1 x\left(-\frac{\pi^2 x^3}{24} - \frac{\pi^4 x^5}{2880} - \cdots\right) dx.$$

In particular, the first integral in the final term can be explicitly evaluated. The

Table 5.1: Estimates of the Mahler measure in lemma 5.9

N	Estimate
1	-0.7354974750452893100066784608239737975163908036
2	-1.8420265499103943019417477652626591228410941526
3	-0.4102009255386977164806383140251091058050097051
4	0.3507832225993429668000625314558852428743006880
5	0.2411625552520558032990599130831707107851139578 6
6	0.2454285734602411467084228017821959082587123861
7	0.2453485703123265911781960631735559649468896105
8	0.2453493068965619254053039477782658300260352989
9	0.2453492985426802975386386596233168514006422190
10	0.2453492984951657022189643319149448066151938057
11	0.2453492984957029003119488640811930734755527995
12	0.2453492984957087695689561142212925829220767722
13	0.2453492984957087697191850787435859789182293099
14	0.2453492984957087696802277722143756962564698188

second integral in the final term can be evaluated using the periodic point expansion and equation 5.3.

Table 5.2: Estimates of the integral in lemma 5.10

N	Estimate
1	0
2	0.0710058106659420626139608915475710754171537460
3	-0.0674079942679211711371277070046100911840936746
4	-0.1137418084423140487912693319297018815707912345
5	-0.1091292853392956782029628035362785456971448144
6	-0.1092261716068447267310745709643933696618902089
7	-0.1092257521118177154674761773445224803026556852
8	-0.1092257434238678523468850000209060091707199826
9	-0.1092257435162616483165814424580769513185926387
10	-0.1092257435159467354031492882515473442102548640

5.6 Related Integrals: Catalan Constants

The problem of evaluating Mahler measures has parallels to evaluating other explicit integrals. A particularly interesting class of such integrals are the following:

$$c_n = \frac{1}{(2\pi)^n} \int_0^{2\pi} \cdots \int_0^{2\pi} \log(2 - 2\cos(\theta_1)) \cdots (2 - 2\cos(\theta_n)) d\theta_1 \cdots d\theta_n,$$

called the *Catalan constants*. These particular integrals occur as limits of determinants of a discrete Laplacian associated to a graph called a discrete torus, given with a suitable normalisation. They are also related to Kirckhoff's 1847 paper on spanning trees.

Example 5.16 ($n = 1$)**.** The constant is named in honour of E. C. Catalan (1814-1894). In 1865, Catalan computed c_1 to 9 decimal places. It has now been computed to over 30×10^9 decimal places by Yee (62). It has many expressions in terms of series, for example the slowly converging series

$$c_1 = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^2}.$$

From our point of view, the most useful form of the constant is as an integral

$$c_1 = -\int_0^{\pi/2} \log(2\sin(t/2))dt = 0.915965\ldots.$$

We can change variables by $t = \frac{\pi}{2}x$, then we can write

$$c_1 = -\frac{\pi}{2} \int_0^1 \log\left(\sin(\pi x/4)\right) dx - \frac{\pi}{2} \log 2.$$

We can expand

$$\begin{aligned}\sin(\pi x/4) &= (\pi x/4) - \frac{(\pi x/4)^3}{3!} + \frac{(\pi x/4)^5}{5!} - \frac{(\pi x/4)^7}{7!} + \cdots \\ &= (\pi x/4)\left(1 - \frac{(\pi x/4)^2}{3!} + \frac{(\pi x/4)^4}{5!} - \frac{(\pi x/4)^6}{7!} + \cdots\right) \\ &= (\pi x/4) f(x)\end{aligned}$$

where

$$f(x) = 1 - \frac{(\pi x/4)^2}{3!} + \frac{(\pi x/4)^4}{5!} - \frac{(\pi x/4)^6}{7!} + \cdots .$$

which is analytic and non-zero on a neighbourhood of the interval $[0, 1]$. We can then write

$$c_1 = -\frac{\pi}{2}\left(\log(\pi/4) + \underbrace{\int_0^1 \log x dx}_{=1} + \int_0^1 \log f(x) dx + \log(2)\right).$$

To estimate c_1, it remains to estimate $\int_0^1 \log f(x) dx$, for which we can use the periodic point expansion described earlier.

Table 5.3: Estimates of the integral in example 5.16

N	Estimate
1	0
2	0.0287834811325220051958276405284368315928709719807 61
3	-0.017597596938717325443892198755114160250026672339315
4	-0.037035295847928556939358308020139133203154818774092
5	-0.034628688321030893474631521272671044247185943206331
6	-0.034705707340429574152204007920956903583981142752151
7	-0.034704504761600623768703825770410214250300363705901
8	-0.034704513378801624771335583022987871957717300165660
9	-0.034704513351055271758860768558861591782522325683191
10	-0.034704513351092439495683762246807986869029078814615

Whereas this may not be the best numerical approach to computation, it does give another series expression for the Catalan constant.

5.7 Applications

5.7.1 Entropy of Dynamical Systems

The Mahler measure plays an important role in the entropy of higher rank actions. In the study of $\mathbb{Z}^d$- subshifts of finite type it is frequently the case that the entropy of the shift can be expressed in terms of a Mahler measure, and thus cannot be explicitly computed. In particular, in a number of interesting cases the entropy can be written explicitly in terms of integrals, thus reducing the estimation of the entropy in these cases to the estimation of these integrals.

In general, an algebraic $\mathbb{Z}^d$-action on a compact abelian group X is a homeomorphism $\alpha : \mathbb{Z}^d \to \mathrm{Aut}(X)$ to the automorphisms $\mathrm{Aut}(X)$ of X.

Let us specialise this to cyclic $\mathbb{Z}^d$-actions, and give the standard setup. Let $\mathbb{Z}[z_1^{\pm 1}, \ldots, z_d^{\pm 1}]$ be the ring of Laurent polynomials with $\mathbb{Z}$-coefficients in the variables $z_1, \ldots, z_d$, i.e., a polynomial $P(\cdot) \in \mathbb{Z}[z_1^{\pm 1}, \ldots, z_d^{\pm 1}]$ can be written in the form

$$P(z_1, \ldots, z_d) = \sum_{-N \le i_1, \cdots, i_d \le N} a_{i_1, \cdots, i_d} z_1^{i_1} \cdots z_d^{i_d}.$$

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the torus then we can define a natural $\mathbb{Z}^d$-action $\sigma^{(m_1, \cdots, m_d)} : \mathbb{T}^{\mathbb{Z}^d} \to \mathbb{T}^{\mathbb{Z}^d}$ on infinitely many products of the torus by

$$\sigma^{(m_1, \cdots, m_d)}(x_{n_1, \cdots, n_d}) = x_{n_1 + m_1, \cdots, n_d + m_d}$$

where $x \in \mathbb{T}^{\mathbb{Z}^d}$, i.e. $(x_{n_1, \ldots, n_d})_{n_1, \ldots, n_d \in \mathbb{Z}} \in \mathbb{T}$. Moreover, we can associate to $P(z_1, \ldots, z_d)$ an action

$$\sum_{-N \le i_1, \cdots, i_d \le N} a_{i_1, \cdots, i_d} \sigma^{(i_1, \cdots, i_d)} : \mathbb{T}^{\mathbb{Z}^d} \to \mathbb{T}^{\mathbb{Z}^d}.$$

We can identify the ring $\mathbb{Z}[z_1^{\pm 1}, \dots, z_d^{\pm 1}]$ with the dual group of $\mathbb{T}^{\mathbb{Z}^d}$ by

$$\langle P, x\rangle = \exp\left(2\pi i \sum_{-N\le i_1,\cdots,i_d\le N} a_{i_1,\cdots,i_d} x_{i_1,\cdots,i_d}\right)$$

where $x \in \mathbb{T}^{\mathbb{Z}^d}$ and $P \in \mathbb{Z}[z_1^{\pm 1}, \dots, z_d^{\pm 1}]$. In particular, the shift map on $\mathbb{T}^{\mathbb{Z}^d}$ corresponds to multiplication by polynomials on $\mathbb{Z}[z_1^{\pm 1}, \dots, z_d^{\pm 1}]$.

A subshift $X \subset \mathbb{T}^{\mathbb{Z}^d}$ is a closed subgroup which is shift invariant. This latter property corresponds to the algebraic property that its annihilator

$$\text{Annih}(X) = \{P \in \mathbb{Z}[z_1^{\pm 1}, \dots, z_d^{\pm 1}] : \langle P, x\rangle = 1 \ \forall x \in X\}$$

in an ideal in $\mathbb{Z}[z_1^{\pm 1}, \cdots, z_d^{\pm 1}]$. Conversely, if $A \subset \mathbb{Z}[z_1^{\pm 1}, \dots, z_d^{\pm 1}]$ is an ideal then we can associate a closed shift invariant space

$$X_A = \{x \in \mathbb{T}^{\mathbb{Z}^d} : \langle P, x\rangle = 1 \ \forall P \in A\}.$$

i.e., the closed shift invariant subgroup annihilated by A. These are the cyclic algebraic $\mathbb{Z}^d$-actions.

Given a Laurent polynomial in d-variables $P(z_1, \cdots, z_d) \in \mathbb{Z}[z_1^{\pm 1}, \cdots, z_d^{\pm 1}]$ we can associate the ideal $A = P(z_1, \cdots, z_d)\mathbb{Z}[z_1^{\pm 1}, \cdots, z_d^{\pm 1}]$. It is these $\mathbb{Z}^d$-actions whose entropy is described by the Mahler measure.

Example 5.17 ($d = 1$)**.** If $P(z) = a_d z^d + \cdots + a_0 \in \mathbb{Z}[z]$ we have that

$$X_P = \{(x_k) \in \mathbb{T}^{\mathbb{Z}} : a_0 x_k + a_1 x_{k+1} + \cdots + a_d z_{k+d} = 0, \forall k\}.$$

One can compute the entropy of such Z^d-actions defined in terms of a polynomial P via its Mahler measure.

Theorem 5.18 (Lind, Schmidt and Ward, (29))**.** *The topological entropy of X_P is* *given my $m(P)$ if P is a prime ideal (and zero otherwise).*

Example 5.19. Let $f(u,v) = 1 + u + v$ be a Laurent polynomial. The associated dynamical space is the set

$$\{x \in \mathbb{T}^{\mathbb{Z}^2} : x_{i,j} + x_{i+1,j} + x_{i,j+1} = 0, \ \forall i, j \in \mathbb{Z}\}.$$

In this case the entropy of the shift is the logarithmic Mahler measure

$$m(f) = \frac{3\sqrt{3}}{4\pi} L(2, \chi_3) = 0.3230659472\ldots.$$

Example 5.20. The same analysis can be used to compute the entropy in a number of examples, and a table is given below from the article (29) of Lind, Schmidt and Ward, which is based on computations of Smyth.

Polynomials	Entropy
$u + v \pm 1$	$\frac{3\sqrt{3}}{4\pi} L(2, \chi_3)$
$u + v + w \pm 1$	$\frac{7\zeta(3)}{2\pi^2}$
$u + v \pm k$	$\log \lvert k \rvert, \lvert k \rvert \geq 2$
$(u+v)^2 \pm 2$	$\frac{1}{2}\log 2 + \frac{2}{\pi} L(2, \chi_4)$
$(u+v)^2 \pm 3$	$\frac{2}{3}\log 3 + \frac{\sqrt{3}}{\pi} L(2, \chi_3)$
$u^2 - v^2 + uv + 3u - v + 1$	$2 \log \rho$

where $L(s, \chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s}$, χ is a Dirichlet character, and $\rho = \frac{1+\sqrt{5}}{2}$.

5.7.2 Volumes of Hyperbolic Manifolds

We can also consider functions closely related to the volume of ideal tetrahedra in three dimensional hyperbolic space. These were studied by Milnor (34), who established the basic formula for their volumes.

Consider the upper half-space model for hyperbolic space

$$\mathbb{H}^3 = \{(x+iy,t) \in \widehat{\mathbb{C}} : t > 0\}$$

with the Poincaré metric

$$ds^2 = \frac{dx^2+dy^2+dt^2}{t^2}.$$

An ideal tetrahedron corresponds to choosing 4 points (vertices) on the unit sphere and joining these with geodesics (in the Poincaré metric) which correspond to the edges of a tetrahedron. One can chose the vertices so that this volume has its largest volume $3L(\pi/3)$ where

$$L(\alpha) = -\int_0^\alpha \log|2\sin(u)|du.$$

This comes from the following characterization of the volume of the tetrahedron.

Lemma 5.21 (Lobachevsky, Milnor (34))**.** *If the three faces meeting at a common vertex z have angles α, β, δ between them, then the volume $D(z)$ is given by $D(z) = L(\alpha) + L(\beta) + L(\delta)$.*

If we add and subtract $\int_0^\alpha \log(2u)du$ then we get:

$$L(\alpha) = \int_0^\alpha \log|2\sin(u)/u|du + \int_0^\alpha \log(2u)du.$$

The first term can be estimated as before with $\alpha = \pi/3$ using periodic points and the second term can be integrated directly.

The volumes of hyperbolic tetrahedra play a role in the values of certain zeta functions. There are many proofs that $\zeta(2) = \sum_{n=1}^\infty \frac{1}{n^2} = \frac{\pi^2}{6}$. More generally, we can look at particular values of the Dedekind zeta function $\zeta_K(s)$, which is the analogue of $\zeta(s)$ for algebraic number fields $K \supset \mathbb{Q}$.

Definition 5.22. For $s \in \mathbb{C}$ we define

$$\zeta_K(s) = \prod_{P \subset \mathcal{O}_K} \left(1 - N_{K/\mathbb{Q}}(P)^{-s}\right)^{-1}$$

where the product is over all prime ideals P of the ring of integers $\mathcal{O}_K$ of K and $N_{K/\mathbb{Q}}(P)$ is its norm.

In particular, $\zeta_K(s)$ converges for $Re(s) > 1$ and has a meromorphic extension to $\mathbb{C}$.

Theorem 5.23 (Zagier, (63))**.** *Fix an algebraic number field K. The value $\zeta_K(2)$ of the Dedekind zeta function is related to volumes of hyperbolic tetrahedra.*

For a specific example of this, we have the following.

Example 5.24. When $K = \mathbb{Q}(\sqrt{-7})$ then the Dedekind zeta function takes the form

$$\zeta_{\mathbb{Q}(\sqrt{-7})}(s) = \frac{1}{2} \sum_{(x,y)\neq(0,0)} \frac{1}{(x^2 + 2xy + 2y^2)^s}$$

and

$$\zeta_{\mathbb{Q}(\sqrt{-7})}(2) = \frac{4\pi^2}{21\sqrt{7}} \left(2D\left(\frac{1+\sqrt{-7}}{2}\right) + D\left(\frac{-1+\sqrt{-7}}{2}\right)\right)$$

which numerically is approximately $1.1519254705\ldots$.

A related value is the dilogarithm.

Definition 5.25. For $m \in \mathbb{N}$, the polylogarithm is defined by

$$Li_m(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^m},$$

for $z \notin [1, \infty)$, the cut line.

In particular, when $m = 2$, the function is called the dilogarithm, and we have

$$Li_2(z) = -\int_0^z \frac{\log|1-u|}{u} du.$$

There are many identities and functional relations for the dilogarithm. Here it is related to the function $D(z)$ by

$$D(z) = \mathfrak{Im}\, Li_2(z) + \arg(1-z)\log|z|.$$

5.7.3 Knot theory and geometry

A section of the excellent survey article of Smyth (59) references some results on the connection between knot theory and Mahler measures, where the Mahler measure of one-variable polynomials arises in connection with Alexander polynomials of knots and reduced Alexander polynomials of links.

Indeed, in Reidemeister's classic book a specific polynomial appears as the Alexander polynomial of a knot. Hironaka has shown that among a wide class of Alexander polynomials of pretzel links, this one has the smallest Mahler measure. Champanerkar and Kofman study a sequence of Mahler measures of Jones polynomials of hyperbolic links. They show that it converges to the Mahler measure of a 2-variable polynomial.

Chapter 6

Entropy Rates of Hidden Markov Processes

6.1 Hidden Markov Chains

Let $Y = \{Y_n\}_n$ denote a process on states $\{1, 2, \cdots, k\}$ and let $Z = \{Z_n\}_n$ denote a process on states $\{1, 2, \cdots, m\}$ (with $m < k$). We assume that there is a Markov probability measure μ_Y on Y, associated to a stochastic matrix P.

Consider a map on states: $\Psi : \{1, 2, \cdots, k\} \to \{1, 2, \cdots, m\}$ and a corresponding map $\Phi(Y) = Z$ where $Z_n = \Psi(Y_n)$. We can consider the factor measure $\mu_Z = \Phi(\mu_X)$, and then the associated measure is a Hidden Markov Chain.

An alternative point of view from digital communications is to describe Z as the output process obtained when passing a finite state Markov chain through a noisy channel. Thus Z can be thought of as the output signal changed with small probability by a Bernoulli 'noise', see the paper of Holliday, Goldsmith, and Glynn (19) for more details.

Simple Example (Symmetric Binary Channel). Let $\{X_n\}$ be a Markov process with associated stochastic matrix

$$\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}.$$

This can be viewed as the 'input'.

At time n a binary symmetric channel with crossover probability p can be characterized by the equation $Z_n = X_n \oplus E_n$, where $X_n \in \{0,1\}$ denotes the binary input, $E_n \in \{0,1\}$ the i.i.d. binary noise with probability vector $(1-\epsilon, \epsilon)$ and $\oplus$ denotes addition modulo 2. Thus $Z = \{Z_n\}$ represents the corrupted output.

More precisely, the product $Y = X \times E$ with $Y_n = (X_n, E_n)$ is again Markov with associated stochastic matrix

$$\begin{pmatrix} p(1-\epsilon) & p\epsilon & (1-p)(1-\epsilon) & (1-p)\epsilon \\ p(1-\epsilon) & p\epsilon & (1-p)(1-\epsilon) & (1-p)\epsilon \\ (1-p)(1-\epsilon) & (1-p)\epsilon & p(1-\epsilon) & p\epsilon \\ (1-p)(1-\epsilon) & (1-p)\epsilon & p(1-\epsilon) & p\epsilon \end{pmatrix}$$

(where $k = 4$ and $m = 2$).

Thus Z is a hidden Markov chain with $Z = \Phi(Y)$ where $\Psi(0,0) = \Psi(1,1) = 0$ and $\Psi(0,1) = \Psi(1,0) = 1$.

Definition 6.1. The entropy rate $Y = \{Y_n\}$ of Y is defined by $H(Y) = \lim_{n\to+\infty} \frac{1}{n} H_n(Y)$, where

$$H_n(Y) = -\sum_{y_0,\cdots,y_{n-1}} \mu_Y[y_0,\cdots,y_{n-1}] \log \mu_Y[y_0,\cdots,y_{n-1}]$$

If μ_Y is a Markov measure corresponding to a stochastic matrix $P = P_Y$, with right eigenvector p, then the entropy is well known to be:

$$H(Y) = -\sum_{i,j} p_i P_{ij} \log P_{ij}.$$

6.2 Statement of Results

We want to employ the work of Han and Marcus (15) in formulating the entropy rate in terms of actions of projective maps on the simplex, so that we can then determine an expression for the entropy rate in terms of the determinant of a transfer operator.

Let $\Lambda = \{x \in \mathbb{R}^d : x_1 + \cdots + x_d = 1,\, x_1, \ldots, x_d > 0\}$, the $(d-1)$-dimensional simplex. Next let $A_1, \ldots, A_d$ be strictly positive $d \times d$ matrices, and let $c_1, \ldots, c_d \in \Lambda$ such that $\sum_k c_k = 1 \in \mathbb{R}^d$. For brevity we define $\kappa(x) = x \log x, \kappa(0) = 0$. The action of the matrix A_j on the simplex Λ is given by the map $F_j : \Lambda \to \Lambda$ defined by

$$F_j(x) = \frac{A_j x}{A_j x \cdot 1},$$

where the denominator normalises $A_j x$ so it is in the simplex again. We also require the maps $r_j : \Lambda \to [0, 1]$ given by

$$r_j(x) = c_j \cdot x.$$

The Blackwell probability measure μ defined on Λ is completely determined by the implicit relation

$$\mu(A) = \sum_{j=1}^{d} \int_{F_j^{-1}(A)} r_j(x) d\mu(x). \tag{6.1}$$

We would like to find a way to efficiently calculate the entropy rate, which is given by the following result of Han and Marcus (15).

Theorem 6.2. *The entropy rate is given by*

$$H = -\int_{\Lambda} \sum_{j=1}^{d} \kappa(r_j(x)) d\mu(x). \tag{6.2}$$

To give an explicit formula for the entropy rate, we need some more definitions. Define $I = \{1, \ldots, d\}$, and for $\underline{i} \in I^n$ for some $n \in \mathbb{N}$, define

- the matrix product $A_{\underline{i}} = A_{i_1} A_{i_2} \cdots A_{i_n}$,
- the function $F_{\underline{i}} = F_{i_1} \circ \cdots \circ F_{i_n}$,
- $X_{\underline{i}} \in \Lambda$ is the fixed point of $F_{\underline{i}}$, and finally
- $\lambda_{\underline{i}} > 0$ is the maximal positive eigenvalue of $A_{\underline{i}}$. We also denote the other eigenvalues by $\lambda_{\underline{i}}^{(k)}$, with $\lambda_{\underline{i}} = \lambda_{\underline{i}}^{(1)}$ and $|\lambda_{\underline{i}}^{(1)}| > |\lambda_{\underline{i}}^{(2)}| \geq \ldots \geq |\lambda_{\underline{i}}^{(d)}|$.

The fixed point $X_{\underline{i}}$ is in fact the eigenvector associated with $\lambda_{\underline{i}}$, suitably scaled so that $X_{\underline{i}} \in \Lambda$. For $1 \leq j, k \leq d$, any $t \in \mathbb{C}$, and any $x \in \Lambda$, define the weight function $W_k^{j,t} : \Lambda \to \mathbb{R}$ by

$$W_k^{j,t}(x) = r_k(x) e^{-t\kappa(r_j(F_k x))}.$$

We extend this definition for tuples. Fix $\underline{i} \in I^n$. We can denote

$$W_{\underline{i}}^{j,t}(x) = W_{i_1}^{j,t}(F_{i_2} \cdots F_{i_n} x) W_{i_2}^{j,t}(F_{i_3} \cdots F_{i_n} x) \cdots W_{i_{n-1}}^{j,t}(F_{i_n} x) W_{i_n}^{j,t}(x), \tag{6.3}$$

where $x \in \Lambda$ and $1 \leq j \leq d$.

Definition 6.3. Define the function $\Delta_j(z,t)$ as follows,

$$\Delta_j(z,t) = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \sum_{\underline{i} \in I^n} \frac{W_{\underline{i}}^{j,t}(X_{\underline{i}})}{\prod_{k=2}^{d} (1 - \lambda_{\underline{i}}^{(k)} / \lambda_{\underline{i}}^{(1)})} \right),$$

for $z \in \mathbb{C}$ and $t \in \mathbb{R}$.

The next proposition is the key technical result which provides the basis for a practical method to calculate the entropy rate numerically.

Proposition 6.4. *We can express the entropy rate H in terms of the function*

$\Delta_j(z,t)$ *using the following formula:*

$$H = \sum_{j=1}^{d} \left(\frac{\partial \Delta_j}{\partial t} \Big/ \frac{\partial \Delta_j}{\partial z} \right) \Bigg|_{z=1,\, t=0}.$$

We can use this to calculate terms in a sequence which converges super-exponentially to H, which are easily calculable from eigenvectors and eigenvalues of the matrices $A_{\underline{i}}$.

We can write

$$\Delta_j(z,t) = 1 + \sum_{n=1}^{\infty} a_{n,j}(t) z^n,$$

for any $z \in \mathbb{C}$ and $t \in \mathbb{R}$, where

$$b_{n,j}(t) = \sum_{\substack{n_1,\dots,n_m \in \mathbb{N} \\ n_1+\cdots+n_m=n}} \frac{(-1)^m}{m!} \prod_{i=1}^{m} \left(\frac{1}{n_i} \sum_{\underline{i} \in I^{n_i}} \frac{W_{\underline{i}}^{j,t}(X_{\underline{i}})}{\prod_{k=2}^{d} (1 - \lambda_{\underline{i}}^{(k)} / \lambda_{\underline{i}}^{(1)})} \right). \tag{6.4}$$

Our main result is the following theorem.

Theorem 6.5. *Define*

$$H_N = \sum_{j=1}^{d} \frac{\sum_{n=1}^{N} \frac{db_{n,j}}{dt}(0)}{\sum_{n=1}^{N} n b_{n,j}(0)}.$$

Then $H_N \to H$ *with*

$$|H - H_N| = \mathcal{O}(\delta^{N^{1+1/d}})$$

for some $0 < \delta < 1$.

6.3 Proofs of Results

As with the chapter on continuous Lyapunov exponents, we would like to move everything to a setting where we can define a nuclear transfer operator. In the continuous Lyapunov exponents chapter, we defined a manifold M (the positive quadrant of complex projective space) and a domain D in $\mathbb{C}^{d-1}$ which is the bijective image of the manifold under its sole atlas map $\phi : M \to D$. We then looked at linear fractional transformations on D which reproduce the action of a matrix on M. Here we say the linear fractional map $T_j : D \to D$ corresponds to the matrix A_j.

We need to find a connection between the simplex Λ and the domain D. Define the complex simplex by

$$\Lambda_{\mathbb{C}} = \left\{ z \in \mathbb{C}^d : \sum_{j=1}^{d} z_j = 1 + i, \mathfrak{Re}\, z_j > 0, \mathfrak{Im}\, z_j > 0 \right\}.$$

This is just two copies of the simplex, and can be thought of as $\Lambda_{\mathbb{C}} = \Lambda + i\Lambda$. We can embed Λ in $\Lambda_{\mathbb{C}}$ using the map $\iota : \Lambda \to \Lambda_{\mathbb{C}}$ defined by

$$\iota(x) = (x_1 + ix_1, x_2 + ix_2, \ldots, x_d + ix_d) = (1+i)x,$$

for $x \in \Lambda$. This maps Λ into the diagonal of $\Lambda_{\mathbb{C}}$, thus we only have an inverse for ι when the real and imaginary parts of each component are equal. We can relate $\Lambda_{\mathbb{C}}$ to M by defining the map $\psi : \Lambda_{\mathbb{C}} \to M$ by $\psi(x) = [x]$, so $\phi(x)$ is the complex line through the origin which contains x. For the inverse, start by defining the scaling factor $S : \mathbb{C}^d \to \mathbb{C}$

$$S(z_1, \ldots, z_d) = \frac{1}{1+i} \sum_{j=1}^{d} z_j,$$

and note that $S(\lambda z_1, \ldots, \lambda z_d) = \lambda S(z_1, \ldots, z_d)$ for any $0 \neq \lambda \in \mathbb{C}$. Now

$$\psi^{-1}[z_1, \ldots, z_d] = (z_1, \ldots, z_d)/S(z_1, \ldots, z_d)$$

for $[z_1, \ldots, z_d] \in M$. This chooses the correct point in a line so that point is also contained in the simplex $\Lambda_{\mathbb{C}}$. This inverse is well-defined since

$$\psi^{-1}[\lambda z_1, \ldots, \lambda z_d] = \lambda(z_1, \ldots, z_d)/(\lambda S(z_1, \ldots, z_d)) = \psi^{-1}[z_1, \ldots, z_d]$$

and $\psi^{-1}[z_1, \ldots, z_d] \in \Lambda_{\mathbb{C}}$ because

$$\frac{1}{S(z_1, \ldots, z_d)} \sum_{j=1}^{d} z_j = 1 + i,$$

and the real and imaginary parts are all strictly positive. We can move between $\Lambda_{\mathbb{C}}$ and D using the map $\omega : \Lambda_{\mathbb{C}} \to D$ given by $\omega = \phi \circ \psi$.

What does the image of the embedding of Λ in $\Lambda_{\mathbb{C}}$ look like under ω? Let $x \in \Lambda$. Then

$$\begin{aligned} \omega(\iota(x)) &= \phi([x_1 + ix_1, \ldots, x_d + ix_d]) \\ &= \left(\frac{x_2 + ix_2}{x_1 + ix_1}, \frac{x_3 + ix_3}{x_1 + ix_1}, \ldots, \frac{x_d + ix_d}{x_1 + ix_1}\right) \\ &= \left(\frac{x_2}{x_1}, \frac{x_3}{x_1}, \ldots, \frac{x_d}{x_1}\right), \end{aligned}$$

which is in $\mathbb{R}_+^{d-1}$, hence $\omega(\iota(\Lambda)) \subset \mathbb{R}_+^{d-1}$. For $x \in \mathbb{R}_+^{d-1}$ we have

$$\begin{aligned} \omega^{-1}(x) &= \psi^{-1}(\phi^{-1}(x)) = \psi^{-1}([1, x_1, \ldots, x_{d-1}]) \\ &= (1, x_1, \ldots, x_{d-1})/S(1, x_1, \ldots, x_{d-1}) \\ &= (1 + i)(1, x_1, \ldots, x_{d-1})/(1 + x_1 + \ldots + x_{d-1}) \end{aligned}$$

and from this we can see the real and imaginary parts of each component are equal, so the inverse of ι is defined, hence $\omega^{-1}(x)$ corresponds to the point

$$(1, x_1, \ldots, x_{d-1})/(1 + x_1 + \ldots + x_{d-1}) \in \Lambda,$$

so $\omega^{-1}(\mathbb{R}_+^{d-1}) \subset \iota(\Lambda)$ and $\omega(\iota(\Lambda)) = \mathbb{R}_+^{d-1}$.

The maps F_j are $\Lambda \to \Lambda$, but we want to extend them to $\Lambda_{\mathbb{C}}$. Define $G_j : \Lambda_{\mathbb{C}} \to \Lambda_{\mathbb{C}}$ by

$$G_i = \psi^{-1} \circ A_j \circ \psi$$

where A_j acts on M, i.e. $A_j[z_1, \ldots, z_d] = [A(z_1, \ldots, z_d)]$. Now we would like that $G_j \circ \iota = \iota \circ F_j$, i.e. G_j does on the embedding of Λ the same thing as F_j does on Λ. Let $x \in \Lambda$. We have

$$\begin{aligned} G_j(\iota(x)) &= \psi^{-1}[\iota(A_j(x))] = \iota(A_j(x))/S(\iota(A_j(x))) \\ &= (1+i)A_j(x)/(S((1+i)A_j(x))) = \iota(A_j(x)/(A_j(x) \cdot 1)) \\ &= \iota(F_j(x)), \end{aligned}$$

as hoped. The map $\omega \circ G_j \circ \omega^{-1} : D \to D$ takes the action of G_j to D. Expanding this we have

$$\begin{aligned} \omega \circ G_j \circ \omega^{-1} &= (\phi \circ \psi) \circ (\psi^{-1} \circ A_j \circ \psi) \circ (\psi^{-1} \circ \phi^{-1}) \\ &= \phi \circ A_j \circ \phi^{-1} = T_j, \end{aligned}$$

so we have our usual complex linear fractional transformation T_j associated with the matrix A_j. We also need to extend the maps $r_j : \Lambda \to [0,1]$. Define $\tilde{r}_j : \Lambda_{\mathbb{C}} \to \mathbb{C}$ by

$$\tilde{r}_j(z) = \frac{z \cdot c_j}{1+i},$$

for $z \in \Lambda_{\mathbb{C}}$, so for $x \in \Lambda$, we have $\tilde{r}_j(\iota(x)) = c_j \cdot x$, i.e. $\tilde{r}_j \circ \iota = r_j$. Define $\gamma_j : D \to \mathbb{C}$ by $\gamma_j = \tilde{r}_j \circ \omega^{-1}$, so for $z \in D$, we have

$$\begin{aligned} \gamma_j(z) &= \tilde{r}_j(\psi^{-1}[1, z_1, \ldots, z_{d-1}]) \\ &= \frac{(1, z_1, \ldots, z_{d-1}) \cdot c_j}{(1, z_1, \ldots, z_{d-1}) \cdot 1}. \end{aligned}$$

This expression looks like a single component of a linear fractional transformation, so because they map $D \to D$, we must have $\mathfrak{Re}\,\gamma_j > 0$. We can get more detailed information looking at $\tilde{r}_j$, but this is enough to be able to take the complex log of γ_j.

The map $\theta = \omega \circ \iota$ is a bijection $\Lambda \to \mathbb{R}_+^{d-1}$. The pushforward of the Blackwell measure μ onto $\mathbb{R}_+^{d-1}$ is denoted by $\theta_*\mu$, and defined by $(\theta_*\mu)(A) = \mu(\theta^{-1}A)$ for any Borel set $A \subset \mathbb{R}_+^{d-1}$. This gives us $\int_\Lambda f d\mu = \int_{\mathbb{R}_+^{d-1}} f \circ \theta^{-1} d(\theta_*\mu)$ for any Borel measurable $f : \Lambda \to \mathbb{C}$. If $x \in \mathbb{R}_+^{d-1}$, then $\theta^{-1}x \in \Lambda$, and

$$\begin{aligned}\theta^{-1}T_k(x) &= \iota^{-1}\psi^{-1}A_k\phi^{-1}(x) = \iota^{-1}G_k\psi^{-1}\phi^{-1}(x) \\ &= \iota^{-1}G_k\iota\iota^{-1}\omega^{-1}(x) = \iota^{-1}\iota F_k\iota^{-1}\omega^{-1}(x) = F_k\theta^{-1}(x). \qquad (6.5)\end{aligned}$$

A similar calculation gives

$$\gamma_k(x) = r_k\theta^{-1}(x). \qquad (6.6)$$

Equation 6.2 then becomes

$$H = -\int_{\mathbb{R}_+^{d-1}} \sum_{j=1}^{d} \kappa(\gamma_j(x))d(\theta_*\mu)(x). \qquad (6.7)$$

If we take an arbitrary Borel set $A \subset \mathbb{R}_+^{d-1}$ and put $\theta^{-1}A$ in to equation 6.1 then we get

$$\begin{aligned}\mu(\theta^{-1}A) &= \sum_{j=1}^{d}\int_\Lambda \chi_{F_j^{-1}(\theta^{-1}A)}(x)r_j(x)d\mu(x) \\ &= \sum_{j=1}^{d}\int_{\mathbb{R}_+^{d-1}} \chi_{F_j^{-1}(\theta^{-1}A)}(\theta^{-1}(x))r_j(\theta^{-1}(x))d(\theta_*\mu)(x) \\ &= \sum_{j=1}^{d}\int_{\mathbb{R}_+^{d-1}} \chi_A(T_j(x))\gamma_j(x)d(\theta_*\mu)(x),\end{aligned}$$

which is

$$(\theta_*\mu)(A) = \sum_{j=1}^{d}\int_{T_j^{-1}A} \gamma_j(x)d(\theta_*\mu)(x), \qquad (6.8)$$

so everything carries across as expected.

Starting with equation 6.8 then using indicator functions, extending to simple functions, and using an approximation argument, we have, for any measurable f :

$\mathbb{R}_+^{d-1} \to \mathbb{C}$,

$$\int_{\mathbb{R}_+^{d-1}} f d(\theta_*\mu) = \int_{\mathbb{R}_+^{d-1}} \sum_{j=1}^{d} f(T_j x)\gamma_j(x) d(\theta_*\mu)(x), \tag{6.9}$$

and the integrand of the right hand side looks like the definition of a transfer operator, although it needs to be defined on a smaller set.

As before, we find an open bounded connected set U in $\mathbb{C}^{d-1}$ such that

$$\overline{\bigcup_{1\le j\le d} T_j(D)} \subset U \subset \overline{U} \subset D,$$

using lemma 3.5. This implies $\overline{\cup_{j=1}^{d} T_j(U)} \subset U$. We can be quite flexible with this choice because of the next lemma, which works with any such U.

Lemma 6.6. *We have*

$$\operatorname{supp} \theta_*\mu \subset \mathbb{R}_+^{d-1} \cap U.$$

Proof. Let $W \subset \mathbb{R}_+^{d-1}$ be any set such that $W \cap (\mathbb{R}_+^{d-1} \cap U) = \varnothing$, and $g : \mathbb{R}_+^{d-1} \to \mathbb{C}$ any Borel measurable function. Then on putting $f = \chi_W \cdot g$ into equation 6.9, since $\chi_W(T_j x) = 0$ for any $x \in U \cap \mathbb{R}_+^{d-1}$, we see $\int_W g d(\theta_*\mu) = 0$. □

With the 'trick' in 2.13 in mind, for $t \in \mathbb{C}$ and $1 \le j \le d$, define the transfer operator $\mathcal{L}_{j,t} : F(U) \to F(U)$ by

$$\mathcal{L}_{j,t} f(z) = \sum_{k=1}^{d} f(T_k z)\gamma_k(z) e^{-t\kappa(\gamma_j(T_k z))}. \tag{6.10}$$

If we put $t = 0$, we get

$$\mathcal{L}_{j,0} f(z) = \sum_{k=1}^{d} f(T_k z)\gamma_k(z),$$

which doesn't depend on j, and we denote this simpler operator $\mathcal{L}$. We now state and prove the properties of $\mathcal{L}_{j,t}$ that we require.

Proposition 6.7. *The transfer operator has the following properties:*

1. *$\mathcal{L}_{j,t}$ is strongly nuclear for any $1 \le j \le d$ and $t \in \mathbb{C}$,*

2. *we have that* $(\mathcal{L}_{j,t})_{t\in\mathbb{C}}$ *is an analytic family, for any* $1 \leq j \leq d$,

3. 1 *is a maximal simple positive eigenvalue for* $\mathcal{L}$, *and*

4. *there is a unique eigenprojection* μ_U *such that* $\mathcal{L}^\star\mu_U = \mu_U$, *i.e. for any* $f \in F(U)$

$$\int_U \mathcal{L} f d\mu_U = \int_U f d\mu_U.$$

Furthermore μ_U *is chosen so that* $\int_U 1 d\mu_U = 1$.

Proof. To prove the first part, the work of Bandtlow and Jenkinson (5) applies directly to give that the transfer operators are nuclear.

To prove the second part, let $\gamma : [0,1] \to \mathbb{C}$ be a smooth closed curve. Fix $1 \leq j \leq d$ For any $w \in F(U)$ and any $z \in U$ we have

$$\left(\oint_\gamma \mathcal{L}_{j,\gamma(t)}\gamma'(t)dt\right) f(z) = \sum_{i=1}^{d} f(T_i z) \oint_\gamma e^{-\gamma(t)\kappa(\gamma_j(T_k z))}\gamma'(t)dt = 0$$

hence $\oint_\gamma \mathcal{L}_{j,\gamma(t)}\gamma'(t)dt = 0$ and hence by Morera's theorem (see proposition 2.1), $\{\mathcal{L}_{j,t}\}_{t\in\mathbb{C}}$ is an analytic family.

To prove the third part, 1 is clearly an eigenvalue for $\mathcal{L}$. Mayer (32) proves a result which applies here to give that 1 is maximal and simple, using a cone argument.

The final part of the proposition is because 1 is also a maximal positive simple eigenvalue for $\mathcal{L}^\star$, so there exists an eigenprojection μ_U such that $\mathcal{L}^\star\mu_U = \mu_U$. It can also be multiplied by any scalar $\alpha \in \mathbb{C}$, so we may assume that $\int_U 1 d\mu_U = 1$. □

We have the following connection between μ_U and $\theta_*\mu$.

Lemma 6.8. *For any function* $f \in F(U)$, *we have*

$$\int_{\mathbb{R}_+^{d-1}\cap U} f|_{\mathbb{R}_+^{d-1}\cap U} d(\theta_*\mu) = \int_U f d\mu_U.$$

Proof. Equation 6.9 becomes

$$\int_{\mathbb{R}_+^{d-1}\cap U} f|_{\mathbb{R}_+^{d-1}\cap U} d(\theta_*\mu) = \int_{\mathbb{R}_+^{d-1}\cap U} (\mathcal{L}f)|_{\mathbb{R}_+^{d-1}\cap U} d(\theta_*\mu)(x),$$

for any $f \in F(U)$, where we have used the definition of $\mathcal{L}$ and the fact that the support of $\theta_*\mu$ is contained in U. Hence $\phi : F(U) \to \mathbb{C}$ defined by $\phi(f) = \int_{\mathbb{R}_+^{d-1}\cap U} f|_{\mathbb{R}_+^{d-1}\cap U} d(\theta_*\mu)$ gives a linear functional $\phi \in F(U)^\star$ such that $\mathcal{L}^\star\phi = \phi$, but μ_U is the unique such eigenprojection. Hence $\mu_U(f) = \alpha\phi(f)$, for some $\alpha \in \mathbb{C}$. We have $\alpha = 1$ because

$$\phi(1) = \int_{\mathbb{R}_+^{d-1}} 1 d(\theta_*\mu) = \int_\Lambda 1 d\mu = \mu(\Lambda) = 1,$$

using the lemma about the support of $\theta_*\mu$, and the definition of the pushforward measure. But the choice of μ_U is normalised so that $\int_U 1 d\mu_U = 1$. $\square$

For convenience define the function $w_k^{j,t} \in F(U)$ by

$$w_k^{j,t}(z) = \gamma_k(z)e^{-t\kappa(\gamma_j(T_k z))}, \tag{6.11}$$

for $z \in U$.

Because of proposition 6.7, we can apply the perturbation theorem. For $t \in B(0,\epsilon)$, let $h_{j,t}$ and $\lambda_{j,t}$ be the eigenvalues and eigenvectors respectively the perturbation theorem gives us. Now $h_{j,t} \in F(U)$, with $t \mapsto h_{j,t}$ and $t \mapsto \lambda_{j,t}$ being holomorphic maps. Differentiate the expression $\lambda_{j,t}h_{j,t} = \mathcal{L}_{j,t}h_{j,t}$,

$$\begin{aligned}
\left(\frac{\partial\lambda_{j,t}}{\partial t}h_{j,t} + \lambda_{j,t}\frac{\partial h_{j,t}}{\partial t}\right)(z) &= \frac{\partial}{\partial t}\mathcal{L}_{j,t}h_{j,t}(z) \\
&= \sum_{k=1}^{d}\left(\frac{\partial h_{j,t}}{\partial t}(T_k z) - \kappa(\gamma_j(T_k z))h_{j,t}(T_k z)\right) w_k^{j,t}(z) \\
&= \mathcal{L}_{j,t}\left(\frac{\partial h_{j,t}}{\partial t} - \kappa\circ\gamma_j\cdot h_{j,t}\right)(z)
\end{aligned}$$

and put $t = 0$, to get

$$\mathcal{L}\left(\frac{\partial h_{j,t}}{\partial t}|_{t=0} - \kappa \circ \gamma_j\right) = \frac{\partial \lambda_{j,t}}{\partial t}|_{t=0} + \frac{\partial h_{j,t}}{\partial t}|_{t=0},$$

then integrate with respect to μ_U, using the $\mathcal{L}$-invariance of μ_U, to get

$$-\int_U \kappa \circ \gamma_j d\mu_U = \frac{\partial \lambda_{j,t}}{\partial t}|_{t=0}. \tag{6.12}$$

Starting with equation 6.7, applying lemma 6.6, then lemma 6.8, and finally using equation 6.12, we have

$$\begin{aligned} H &= \int_{\mathbb{R}_+^{d-1}} \sum_{j=1}^{d} \kappa(\gamma_j(x)) d(\theta_*\mu)(x) \\ &= \int_{\mathbb{R}_+^{d-1}\cap U} \sum_{j=1}^{d} \kappa(\gamma_j(x)) d(\theta_*\mu)(x) \\ &= \int_U \sum_{j=1}^{d} \kappa(\gamma_j(x)) d\mu_U(x) \\ &= \sum_{j=1}^{d} \frac{\partial \lambda_{j,t}}{\partial t}|_{t=0}. \end{aligned}$$

We summarise this in the following proposition.

Proposition 6.9. *The entropy rate is given in terms of the transfer operator by the formula*

$$H = \sum_{j=1}^{d} \frac{\partial \lambda_{j,t}}{\partial t}|_{t=0}.$$

We can get the eigenvalues from the dynamical zeta function associated with $\mathcal{L}_{j,t}$. This requires formula for the trace of $\mathcal{L}_{j,t}^n$, which requires a formula for $\mathcal{L}_{j,t}^n$. Define $I = \{1, 2, \ldots, d\}$, and for $\underline{i} \in I^n$, we write

$$T_{\underline{i}} = T_{i_1} \circ T_{i_2} \circ \cdots \circ T_{i_n},$$

and define the function $w_{\underline{i}}^{j,t} \in F(U)$ by

$$w_{\underline{i}}^{j,t}(z) = w_{i_1}^{j,t}(T_{i_2}T_{i_3}\cdots T_{i_n}z)w_{i_2}^{j,t}(T_{i_3}\cdots T_{i_n}z)\cdots w_{i_{n-1}}^{j,t}(T_{i_n}z)w_{i_n}^{j,t}(z)$$

where $z \in U$. An inductive argument shows that for any $n \in \mathbb{N}$,

$$\mathcal{L}_{j,t}^n f(z) = \sum_{\underline{i}\in I^n} f(T_{\underline{i}}z)w_{\underline{i}}^{j,t}(z).$$

Because the operators are strongly nuclear, we can define the dynamical determinant $\Delta_j : \mathbb{C}^2 \to \mathbb{C}$ by

$$\Delta_j(z,t) = \det(I - z\mathcal{L}_{j,t}) = \exp\left(-\sum_{n=1}^{\infty}\frac{z^n}{n}\operatorname{tr}\mathcal{L}_{j,t}^n\right), \tag{6.13}$$

for $z \in \mathbb{C}$ and $t \in \mathbb{C}$, which has the property that it is holomorphic separately for small t and any z, and also has the property that, for any t,

$$\Delta_j(\lambda_{j,t}^{-1}, t) = 0,$$

which, by the implicit function theorem, gives

$$0 = \frac{d}{dt}\Delta_j(\lambda_{j,t}^{-1}, t) = -\frac{\partial\Delta_j}{\partial z}(\lambda_{j,t}^{-1}, t)\lambda_{j,t}^{-2}(t)\frac{\partial\lambda_{j,t}}{\partial t}(t) + \frac{\partial\Delta_j}{\partial t}(\lambda_{j,t}^{-1}, t).$$

If we put $t = 0$, use that $\lambda_{j,0} = 1$, and rearrange, we get,

$$\frac{\partial\lambda_{j,t}}{\partial t}(0) = \frac{\partial\Delta_j}{\partial t}(1,0)\bigg/\frac{\partial\Delta_j}{\partial z}(1,0)\,. \tag{6.14}$$

This provides the link between the dynamical determinant and the entropy rate. It remains to obtain a formula for the traces, and to examine the rate of convergence of approximations to the dynamical determinant.

Lemma 6.10. *We have the following formula for the trace, for any $n \in \mathbb{N}$,*

$$\operatorname{tr} \mathcal{L}_{j,t}^n = \sum_{\underline{i} \in I^n} \frac{W_{\underline{i}}^{j,t}(X_{\underline{i}})}{\prod_{k=2}^d (1 - \lambda_{\underline{i}}^{(k)}/\lambda_{\underline{i}}^{(1)})},$$

where $\lambda_{\underline{i}}^{(k)}$ is the k^{th} eigenvalue of $A_{\underline{i}}$, with the maximal positive eigenvalue first and the rest in any order.

Proof. From equation 2.9, the trace in these circumstances is as follows:

$$\operatorname{tr} \mathcal{L}_{j,t}^n = \sum_{\underline{i} \in I^n} \frac{w_{\underline{i}}^{j,t}(x_{\underline{i}})}{\det(I - DT_{\underline{i}}(x_{\underline{i}}))}.$$

For $x \in \mathbb{R}_+^{d-1}$, using equations 6.5 and 6.6 we can write the weight (equation 6.11) $w_k^{j,t}(x)$ as follows

$$w_k^{j,t}(x) = \gamma_k(x) e^{-t\kappa(\gamma_j(T_k x))} = r_k(\theta^{-1}x) e^{-t\kappa(r_j(F_k \theta^{-1} x))}$$

Hence $w_{\underline{i}}^{j,t}(x) = W_{\underline{i}}^{j,t}(\theta^{-1}x)$, where $W_{\underline{i}}^{j,t}$ is defined in equation 6.3. We have that the fixed point for the fractional transformation $T_{\underline{i}}$ corresponds to the fixed point for the action of the matrix on the simplex $F_{\underline{i}}$, i.e. $X_{\underline{i}} = \theta^{-1}x_{\underline{i}}$.

The denominator of the trace formula is due to lemma 3.21. □

Putting the equation for the trace into equation 2.3 gives us the determinant power series expansion coefficients, equation 6.4. We have the following result about the speed of their convergence.

Lemma 6.11. *There exists a constant $0 < \theta < 1$ such that*

$$b_{n,j}(0) = \mathcal{O}(\theta^{n^{1+1/d}}).$$

Proof. Bandtlow and Jenkinson in (3) prove bounds on eigenvalues of a class of transfer operators. The transfer operators here would be in that class, except they prove the result for the transfer operators acting on a different space of functions. However, both their space and the space we work with are 'favourable' spaces, in the sense of another paper of Bandtlow and Jenkinson (5). This means the eigenvalue sequence is the same, so the bounds must also be the same. We then use these bounds on the eigenvales to apply proposition 2.9. □

We can use this result to prove theorem 6.5. First we require a simple lemma.

Lemma 6.12. *If a sequence $(x_n)_{n\in\mathbb{N}}$ has $x_n = \mathcal{O}(\theta^{N^{1+1/d}})$ for some $0<\theta<1$, then*

$$\left|\sum_{n=N+1}^{\infty} x_n\right| = \mathcal{O}(\theta^{N^{1+1/d}}).$$

Proof. Since $(N+1)^{1+1/d} = (N+1)(N+1)^{1/d} > (N+1)N^{1/d} > N^{1+1/d}+1$, we have by induction that $(N+n)^{1+1/d} > N^{1+1/d}+n$, for any $N,n\in\mathbb{N}$. This gives that $\theta^{(N+n)^{1+1/d}} < \theta^{N^{1+1/d}}\theta^n$. Assume N is large enough so that $n \geq N$ implies $|x_n| \leq A\theta^{n^{1+1/d}}$ for some constant $A>0$. Then

$$\left|\sum_{n=N+1}^{\infty} x_n\right| \leq \sum_{n=1}^{\infty}|x_{N+n}| \leq A\sum_{n=1}^{\infty}\theta^{(N+n)^{1+1/d}} \leq A\theta^{N^{1+1/d}}\sum_{n=1}^{\infty}\theta^n = \frac{A\theta}{1-\theta}\theta^{N^{1+1/d}},$$

which proves the lemma. □

Proposition 6.9, equation 6.14, then the power series expansion of Δ_j, give that

$$\begin{aligned} H &= \sum_{j=1}^{d}\frac{\partial\lambda_{j,t}}{\partial t}|_{t=0} \\ &= \sum_{j=1}^{d}\frac{\partial\Delta_j}{\partial t}(1,0)\Big/\frac{\partial\Delta_j}{\partial z}(1,0) \\ &= \sum_{j=1}^{d}\frac{\sum_{n=1}^{\infty}\frac{db_{n,j}}{dt}(0)}{\sum_{n=1}^{\infty}nb_{n,j}(0)}. \end{aligned}$$

So define

$$C_{N,j} = \sum_{n=1}^{N}\frac{db_{n,j}}{dt}(0), \quad D_{N,j} = \sum_{n=1}^{N}nb_{n,j}(0),$$

and

$$H_N = \sum_{j=1} \frac{C_{N,j}}{D_{N,j}},$$

where $N \in \mathbb{N}$. The statement of theorem 6.5 is that there exists a $0 < \delta < 1$ such that $|H - H_N| = \mathcal{O}(\delta^{N^{1+1/d}})$, which is certainly true if we can show this rate of convergence for $|C_{\infty,j}/D_{\infty,j} - C_{N,j}/D_{N,j}|$ for each j, where $C_{\infty,j} = \lim_{N\to\infty} C_{N,j}$ and $D_{\infty,j} = \lim_{N\to\infty} D_{N,j}$. So if we fix j and N, we have

$$\begin{aligned} \left|\frac{C_{\infty,j}}{D_{\infty,j}} - \frac{C_{N,j}}{D_{N,j}}\right| &= \left|\frac{C_{\infty,j}D_{N,j} - C_{N,j}D_{\infty,j}}{D_{N,j}D_{\infty,j}}\right| \\ &\leq \frac{|C_{\infty,j} - C_{N,j}|}{|D_{\infty,j}|} + \frac{|C_{N,j}|}{|D_{\infty,j}|}\frac{|D_{N,j} - D_{\infty,j}|}{|D_{N,j}|}. \end{aligned}$$

We can apply lemma 6.12 to $|C_{\infty,j} - C_{N,j}|$ because $\frac{db_{n,j}}{dt}(0)$ is $\mathcal{O}(\theta^{n^{1+1/d}})$ by Cauchy's estimate. We can also apply the lemma to $|D_{\infty,j} - D_{N,j}|$ because $\mathcal{O}(n\theta^{n^{1+1/d}}) = \mathcal{O}(\theta^{n^{1+1/d}})$. This shows that $|C_{\infty,j}/D_{\infty,j} - C_{N,j}/D_{N,j}| = \mathcal{O}(\theta^{n^{1+1/d}})$ and theorem 6.5 follows.

We are particularly interested in the case for 2×2 matrices, where we can give much more detail for the rate of convergence. If we have

$$A_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix},$$

then from the calculations in example 3.2, we have that the associated linear fractional transformation T_i maps D to the complex Euclidean ball given by

$$B\left(\frac{b_i c_i + a_i d_i}{2a_i b_i}, \frac{|\det A_i|}{2a_i b_i}\right) \subset \mathbb{C}.$$

Choose $x_0 \in \mathbb{R}$ so that $B(x_0, r)$ contains $\cup_{i=1}^d T_i(D)$ and $r > 0$ is minimised. Set $R = x_0$ and define $U = B(x_0, R)$. As before, define the transfer operator on $F(U)$. We have $B(x_0, r) \subset B(x_0, R) \subset D$. This setup minimises the ratio r/R. Using

Cauchy's integral equation (because the functions are bounded, we can extend the domain slightly) and a geometric summation, we can then expand

$$\begin{aligned}
\mathcal{L}_{j,t} f(z) &= \frac{1}{2\pi i}\int_{\partial U} \frac{\mathcal{L}_{j,t} f(\xi)}{\xi - z} d\xi \\
&= \frac{1}{2\pi i}\int_{\partial U} \frac{1}{\xi - x_0}\frac{\mathcal{L}_{j,t} f(\xi)}{1-\frac{z-x_0}{\xi-x_0}} d\xi \\
&= \sum_{n=0}^{\infty}\left(\frac{R^n}{2\pi i}\int_{\partial U}\frac{\mathcal{L}_{j,t} f(\xi) d\xi}{(\xi - x_0)^{n+1}}\right)\left(\frac{z-x_0}{R}\right)^n = \sum_{n=0}^{\infty}\phi_n(f) u_n(z),
\end{aligned}$$

where

$$\phi_n(f) = \frac{R^n}{2\pi i}\int_{\partial U}\frac{\mathcal{L}_{j,t} f(\xi) d\xi}{(\xi - x_0)^{n+1}}, \qquad \text{and} \qquad u_n(z) = R^{-n}(z - x_0)^n.$$

Note that then u_n is normalised. This shows explicitly the nuclear form of the operator. We can estimate $\phi_n(f)$,

$$|\phi_n(f)| \le \frac{R^n}{2\pi}\int_{\partial U}\frac{|\mathcal{L}_{j,t} f(\xi)| d\xi}{R^{n+1}} \le \frac{1/R}{2\pi}\int_{\partial U}|\mathcal{L}_{j,t} f(\xi)| d\xi \le \sup_{\xi\in\partial U}|\mathcal{L}_{j,t} f(\xi)|.$$

Hence, putting $f = u_m$ and using the definition of the transfer operator, we have

$$\begin{aligned}
|\phi_n(u_m)| &\le \sup_{\xi\in\partial U}\sum_{i=1}^{d} R^{-m}|T_i\xi - x_0|^m |w_i^{j,t}(\xi)| \\
&\le (r/R)^m \sup_{\xi\in\partial U}\sum_{i=1}^{d}|w_i^{j,t}(\xi)| = A(r/R)^m
\end{aligned}$$

where $A = \sup_{\xi\in\partial U}\sum_{i=1}^{d}|w_i^{j,t}(\xi)|$. Thus we may use proposition 2.8 to write

$$|b_{n,j}| \le n^{n/2} A^n (r/R)^{n^2} (r/R)^{P(-n\log(r/R))} \tag{6.15}$$

where $P(x) = 1 + x + x^2/2$.

6.4 Examples

Example 6.13. For $d = 2$, let

$$A_1 = \begin{pmatrix} (1-\epsilon)p & (1-\epsilon)(1-p) \\ \epsilon(1-p) & \epsilon p \end{pmatrix}, \text{ and } A_2 = \begin{pmatrix} \epsilon(1-p) & \epsilon p \\ (1-\epsilon)p & (1-\epsilon)(1-p) \end{pmatrix}.$$

Next let

$$\underline{c}_1 = ((1-\epsilon)p + \epsilon(1-p), (1-\epsilon)(1-p) + \epsilon p),$$

$$\underline{c}_2 = (\epsilon p + (1-\epsilon)(1-p), \epsilon(1-p) + (1-\epsilon)p).$$

From (20), this represents a binary symmetric channel with crossover probability p, and noise ϵ. The previously explained approach was used to calculate the entropy rate in Mathematica for $\epsilon = 0.4, p = 0.3$:

n	H_n
1	0.69290250011323568631575581658744020542059956693
2	0.69300443456063581012726571382865227294742708049
3	0.69299982876011041286624242363468938943928590859
4	0.69299990801652614366992392885128419504043082676
5	0.69299990790096214045540311270272917187282066422
6	0.69299990790171531857424334392040229810931367191
7	0.69299990790171539135914580848590178636480965957
8	0.69299990790171539152144927975264738237152074458
9	0.69299990790171539152144403147441929051373903138
10	0.69299990790171539152144403209168962941005432321
11	0.69299990790171539152144403209166373755059831922
12	0.69299990790171539152144403209166373761615259470
13	0.69299990790171539152144403209166373761615154208
14	0.69299990790171539152144403209166373761615154208

This strongly suggests that

$$H \approx 0.69299990790171539152144403209166373761615154208\ldots$$

is accurate to 47 decimal places.

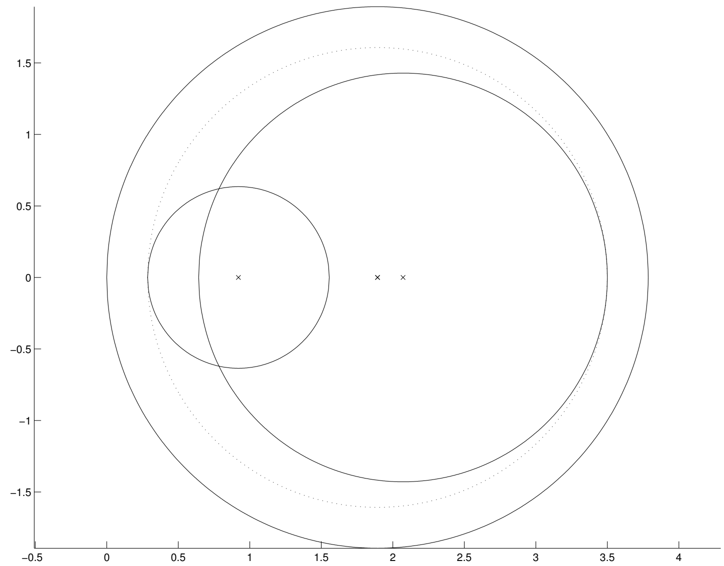


Figure 6.1: The largest circle is $U = B(x_0, R) \subset D$ and the smaller circles are the images of the inverse branches. The dashed circle is $B(x_0, r)$, which bounds the images of the inverse branches.

We now investigate the convergence rate. Figure 6.1 shows a plot of the set U from the previous section, and the images of the inverse branches. We have that $r/R < 0.849056$. We can also make A arbitrarily close to 1.08, by making t arbitrarily small, so we can assume $A = 1.081$ for example. We need a lower bound for $|D_{\infty,j}|$, and by the reverse triangle inequality,

$$|D_{\infty,j}| \geq \left| |D_{N,j}| - \left| \sum_{n=N+1}^{\infty} n b_{n,j}(0) \right| \right| \geq |\,|D_{N,j}| - A_D \theta_D (1-\theta_D)^{-1} \theta_D^{N^2}|$$

providing

$$\left|\sum_{n=N+1}^{\infty} nb_{n,j}(0)\right| \leq A_D\theta_D(1-\theta_D)^{-1}\theta_D^{N^2} \leq |D_{N,j}|.$$

We can use this with equation 6.16, and apply lemma 6.12 to get

$$\left|\frac{C_{\infty,j}}{D_{\infty,j}} - \frac{C_{N,j}}{D_{N,j}}\right| \leq \frac{A_C\theta_C(1-\theta_C)^{-1}\theta_C^{N^2} + |C_{N,j}/D_{N,J}|A_D\theta_D(1-\theta_D)^{-1}\theta_D^{N^2}}{|\,|D_{N,j}| - A_D\theta_D(1-\theta_D)^{-1}\theta_D^{N^2}|} \tag{6.16}$$

where for all $n \geq N$ we have

$$\frac{db_{n,j}}{dt}(0) \leq A_C\theta_C^{N^2}, \quad \text{and} \quad nb_{n,j}(0) \leq A_D\theta_D^{N^2},$$

with $A_C, A_D > 0$, and $0 < \theta_C, \theta_D < 1$. By plotting equation 6.15, using the values $A = 1.081$, $r/R = 0.849056$, $N = 10$ we can experimentally obtain values $A_C = 150$, $A_D = 260$, $\theta_C = 0.91$, and $\theta_D = 0.92$. Equation 6.16 gives $|H - H_{10}| < 0.1446$. Unfortunately this is a very poor bound. If we choose $N = 14$, we can choose $A_C = 1500$, $A_D = 2300$, $\theta_C = 0.91$, and $\theta_D = 0.92$. This gives $|H - H_{14}| < 0.34 \times 10^{-4}$, which is a bigger improvement but the bounds are clearly far from being tight.

Example 6.14. If $\epsilon = 0.45$ and $p = 0.55$ then $r/R < 0.381097$, so we expect faster convergence than the previous example, and we can take $A = 1.011$. We then have the following approximations, which do indeed appear to converge faster.

n	H_n
1	0.693146655725892530369315422823363250823726505498084809878
2	0.693146675708393644778177681175205619060369053186758456456
3	0.693146675701790217741211880682571201377382267567828517623
4	0.693146675701786309857354113570946909149005021361093018604
5	0.693146675701786309868907556388385711579322401084133855863
6	0.693146675701786309868907611177961349014695604159553224450
7	0.693146675701786309868907611177959797280604172163628509822
8	0.693146675701786309868907611177959797280537771413448158947
9	0.693146675701786309868907611177959797280537771413465934661
10	0.693146675701786309868907611177959797280537771413465934661

We can also now experimentally obtain much better bounds. Investigating equation 6.15 shows we can take $A_C = A_D = 0.05$, and $\theta_C = \theta_D = 0.39$ for the values in equation 6.16. This gives $|H - H_{10}| < 0.93 \times 10^{-44}$, which is a much more useful bound.

Bibliography

[1] R. L. Adler and L. Flatto. Geodesic flows, interval maps, and symbolic dynamics. *Bulletin of the AMS*, 25(2), 1991.

[2] V. Baladi. *Positive Transfer Operators and Decay of Corelations*. World Scientific, 2000.

[3] O. Bandtlow and O. Jenkinson. Explicit eigenvalue estimates for transfer operators acting on spaces of holomorphic functions. *Adv. Math.*, 218, no. 3:902–925, 2008.

[4] O. Bandtlow and O. Jenkinson. Invariant measures for real analytic expanding maps. *J. Lond. Math. Soc. (2)*, 75:343–368, 2008.

[5] O. Bandtlow and O. Jenkinson. On the Ruelle eigenvalue sequence. *Ergodic Theory and Dynamical Systems*, 6:1701–1711, 2008.

[6] A. F. Beardon. *The geometry of discrete groups*. Springer-Verlag, Berlin, 1982.

[7] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*. Springer-Verlag, Berlin-New York, 1975.

[8] R. Bowen. Invariant measures for Markov maps of the interval. *Comm. Math. Phys.*, 69 (1):1–17, 1979.

[9] R. Bowen and C. Series. Markov maps associated to Fuchsian groups. *Inst. Hautes Études Sci. Publ. Math*, 50:153–170, 1979.

[10] C. J. Earle and R. S. Hamilton. A fixed point theorem for holomorphic mappings. *Global Analysis, Proc. Symp. Pure Math.*, 16:61–65, 1970.

[11] G. Everest and T. Ward. *Heights of polynomials and entropy in algebraic dynamics*. Springer-Verlag, 1991.

[12] D. Fried. The zeta functions of Ruelle and Selberg. *Ann. Sci. École Norm. Sup.*, 19, no. 4, 1986.

[13] A. Grothendieck. Produits tensoriels topologiques et espaces nucléaires. *Memoirs of the A.M.S.*, 16, 1955.

[14] J. Hadamard. Résolution d'une question relative aux déterminants. *Bull. Sci. Math.*, 28:240–246, 1893.

[15] G. Han and B. Marcus. Derivatives of entropy rate in special families of hidden Markov chains. *IEEE Trans. Inform. Theory*, 53, no. 7:2642–2652, 2007.

[16] S. Helgason. *Topics in harmonic analysis on homogeneous spaces*. Birkhäuser, 1981.

[17] H. Hennion and L. Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*. Springer-Verlag, Berlin, 2001.

[18] F. Hofbauer and G. Keller. Zeta-functions and transfer-operators for piecewise linear transformations. *J. Reine Angew. Math.*, 352:100–113, 1984.

[19] T. Holliday, A. Goldsmith, and P. Glynn. Capacity of finite state channels based on Lyapunov exponents of random matrices. *IEEE Trans. Inform. Theory*, 52, no. 8:3509–3532, 2006.

[20] P. Jacquet, G. Seroussi, and W. Szpankowski. On the entropy of a hidden Markov process. In *Proceedings of the 2004 IEEE International Symposium on Information Theory*, 2004.

[21] O. Jenkinson and M. Pollicott. Computing invariant densities and metric entropy. *Comm. Math. Phys.*, 211:687–703, 2000.

[22] O. Jenkinson and M. Pollicott. Computing the dimension of dynamically defined sets: E_2 and bounded continued fractions. *Ergodic Theory and Dynam. Systems*, 21 no. 5:1429–1445, 2001.

[23] O. Jenkinson and M. Pollicott. Calculating Hausdorff dimensions of Julia sets and Kleinian limit sets. *Amer. J. Math.*, 124, no. 3:495–545, 2002.

[24] O. Jenkinson and M. Pollicott. A dynamical approach to accelerating numerical integration with equidistributed points. *Tr. Mat. Inst. Steklova*, 256:275–289, 2007.

[25] M. Kac. Can one hear the shape of a drum? *American Mathematical Monthly*, 73:1–23, 1966.

[26] T. Katō. *Perturbation Theory for Linear Operators*. Springer, 1995.

[27] S. G. Krantz. *Function theory of several complex variables*. AMS, 1992.

[28] A. Lasota and J. Yorke. On the existence of invariant measures for piecewise monotonic transformations. *Trans. Amer. Math. Soc.*, 186:481–488.

[29] D. Lind, K. Schmidt, and T. Ward. Mahler measure and entropy for commuting automorphisms of compact groups. *Invent. Math.*, 101, no. 3:593–629, 1990.

[30] B. Maskit. On Poincaré's theorem for fundamental polygons. *Advances in Math.*, 7:219–230, 1971.

[31] D. Mayer. On a zeta function related to the continued fraction transformation. *Bull. Soc. Math. France*, 104, no. 2:195–203, 1976.

[32] D. Mayer. Spectral properties of certain composition operators arising in statistical mechanics. *Comm. Math. Phys.*, 68, no. 1:1–8, 1979.

[33] D. H. Mayer. On composition operators on Banach spaces of holomorphic functions. *J. Funct. Anal.*, 35, no. 2:191–206, 1980.

[34] J. Milnor. Hyperbolic geometry: the first 150 years. *Bull. Amer. Math. Soc. (N.S.)*, 6, no. 1:924, 1982.

[35] S. Morita. *Geometry of Differential Forms.* American Mathematical Society, 1997.

[36] R. D. Nussbaum. The radius of the essential spectrum. *Duke Math. J.*, 37:473–478, 1970.

[37] W. Parry. Bowen's equidistribution theory and the Dirichlet density theorem. *Ergodic Theory Dynam. Systems*, 4, no. 1:117–134, 1984.

[38] W. Parry and M. Pollicott. *Zeta functions and the periodic orbit structure of hyperbolic dynamics.* Astérisque, 1990.

[39] A. Pietsche. *Eigenvalues and s-numbers.* Cambridge University Press, 1987.

[40] V. Pit. Invariant relations for the Bowen-Series transform. *Conform. Geom. Dyn.*, 16:103–123, 2012.

[41] M. Pollicott. Meromorphic extensions of generalised zeta functions. *Invent. Math.*, 85, no. 1:147–164, 1986.

[42] M. Pollicott. Some applications of thermodynamic formalism to manifolds with constant negative curvature. *Advances in Mathematics*, 85, 1991.

[43] M. Pollicott. Maximal Lyapunov exponents for random matrix products. *Invent. Math.*, 181, no. 1:209–226, 2010.

[44] M. Pollicott. Computing entropy rates for hidden Markov processes. In *Entropy of hidden Markov processes and connections to dynamical systems*, volume 385, pages 223–245. London Math. Soc. Lecture Note Ser., 2011.

[45] M. Pollicott and A. Rocha. A remarkable formula for the determinant of the Laplacian. *Inventiones Mathematicae*, 130:399–414, 1997.

[46] W. Rudin. *Real and complex analysis.* McGraw-Hill, 1986.

[47] D. Ruelle. Zeta-functions for expanding maps and Anosov flows. *Inventiones Mathematicae*, 34:231–242, 1976.

[48] D. Ruelle. *Thermodynamic Formalism.* Addison-Wesley, 1978.

[49] D. Ruelle. *Dynamical zeta functions for piecewise monotone maps of the interval.* AMS, 1994.

[50] H. H. Rugh. Cones and gauges in complex spaces: Spectral gaps and complex Perron-Frobenius theory. *Annals of Mathematics*, 71:1707–1752, 2010.

[51] R. A. Ryan. *Introduction to tensor products of Banach spaces.* Springer-Verlag, 2002.

[52] O. Sarig. Thermodynamic formalism for countable markov shifts. Lecture notes, `http://www.wisdom.weizmann.ac.il/~sarigo/TDFnotes.pdf`.

[53] H. H. Schaefer. *Topological vector spaces.* Graduate Texts in Mathematics, Vol. 3. Springer-Verlag, 1971.

[54] F. Schweiger. *Multidimensional continued fractions.* Oxford University Press, 2000.

[55] E. Seneta. *Non-negative matrices and Markov chains.* Springer-Verlag, 1973.

[56] Y. G. Sinai. Gibbs measures in ergodic theory. *Uspehi Mat. Nauk 27*, 4(166):21–64, 1972.

[57] C. Smyth. On the product of the conjugates outside the unit circle of an algebraic integer. *Bull. London Math. Soc.*, 3:169–175, 1971.

[58] C. J. Smyth. On measures of polynomials in several variables. *Bull. Australian Math. Soc.*, 23:49–63, 1981.

[59] C. J. Smyth. The mahler measure of algebraic numbers: a survey. In *Number theory and polynomials*, volume 352 of *London Math. Soc. Lecture Note Ser.* Cambridge Univ. Press, Cambridge, 2008.

[60] A. Strohmaier and V. Uski. An algorithm for the computation of eigenvalues, spectral zeta functions and zeta-determinants on hyperbolic surfaces. *Pre-print*, 2012.

[61] P. Walters. *An introduction to ergodic theory.* Springer-Verlag, 1982.

[62] A. J. Yee. Large computations. `http://www.numberworld.org/nagisa_runs/computations.html`.

[63] D. Zagier. Hyperbolic manifolds and special values of Dedekind zeta-functions. *Invent. Math.*, 83, no. 2:285301, 1986.